

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

Yu, Misook

From: Yu, Misook
Sent: Wednesday, September 11, 2002 11:10 AM
To: STIC-Biotech/ChemLib; Chan, Christina
Subject: Request for Rush SEARCH for 09/811,045

This case is due this bi-week.
Please search SEQ ID NO:1 and DNA encoding SEQ ID NO:1.

Chris, please approve the rush search.

Thank you.

Examiner Misook Yu, Ph.D.
703-308-2454 (Phone)
Art Unit 1642
CM1-8E18 (Room)
CM1-8E12 (Mail Box)

*Now teach SEQ ID NO:1 even their own PNAS paper does not
match. 1997 onco gene paper very similar to SEQ ID NO:1
but the date not good.*

THIS PAGE BLANK (USPTO)

STIC-Biotech/ChemLib

75385

Fr m: Chan, Christina
Sent: Wednesday, September 11, 2002 11:25 AM
To: Yu, Misook; STIC-Biotech/ChemLib
Subj ct: RE: Request for Rush SEARCH for 09/811,045

Please rush. Thanks Chris

-----Original Message-----

From: Yu, Misook
Sent: Wednesday, September 11, 2002 11:10 AM
T : STIC-Biotech/ChemLib; Chan, Christina
Subject: Request for Rush SEARCH for 09/811,045

This case is due this bi-week.
Please search SEQ ID NO:1 and DNA encoding SEQ ID NO:1.

Chris, please approve the rush search.

Thank you.

Examiner Misook Yu, Ph.D.
703-308-2454 (Phone)
Art Unit 1642
CM1-8E18 (Room)
CM1-8E12 (Mail Box)

Point of Contact:
Susan Hanley
Technical Info. Specialist
CM1 6B05 Tel: 305-4053

RECEIVED
SEP 11 2002
STIC

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

10/10/2010
10/10/2010
10/10/2010
10/10/2010

THIS PAGE BLANK (USPTO)

09/811,045

WEST Search History

DATE: Thursday, September 12, 2002

| <u>Set Name</u> side by side | <u>Query</u> | <u>Hit Count</u> | <u>Set Name</u> result set |
|---------------------------------|-----------------|------------------|-------------------------------|
| <i>DB=USPT; PLUR=YES; OP=OR</i> | | | |
| L3 | L2 and antibody | 3 | L3 |
| L2 | L1 and hnRNP | 3 | L2 |
| L1 | p2p | 55 | L1 |

END OF SEARCH HISTORY

THIS PAGE BLANK (USPTO)

Your SELECT statement is:
s p2p

| Items | File |
|-------|--|
| 19 | 5: Biosis Previews(R)_1969-2002/Sep W1 |
| 59 | 34: SciSearch(R) Cited Ref Sci_1990-2002/Sep W2 |
| 3 | 35: Dissertation Abs Online_1861-2002/Aug |
| 26 | 65: Inside Conferences_1993-2002/Sep W2 |
| 4 | 71: ELSEVIER BIOBASE_1994-2002/Sep W2 |
| 11 | 73: EMBASE_1974-2002/Aug W4 |
| 47 | 94: JICST-EPlus_1985-2002/Jul W2 |
| 13 | 98: General Sci Abs/Full-Text_1984-2002/Aug |
| 60 | 144: Pascal_1973-2002/Sep W2 |
| 2 | 149: TGG Health&Wellness DB(SM)_1976-2002/Sep W1 |
| 12 | 155: MEDLINE(R)_1966-2002/Sep W2 |
| 3 | 156: ToxFile_1965-2002/Sep W2 |
| 3 | 159: Cancerlit_1975-2002/Jul |
| 1 | 162: CAB HEALTH_1983-2002/Jul |
| 1 | 172: EMBASE Alert_2002/Sep W2 |
| 1 | 266: FEDRIP_2002/Jul |
| 2 | 369: New Scientist_1994-2002/Aug W2 |
| 19 | 399: CA SEARCH(R)_1967-2002/UD=13711 |
| 4 | 434: SciSearch(R) Cited Ref Sci_1974-1989/Dec |
| 1 | 442: AMA Journals_1982-2002/Aug B1 |

SYSTEM:OS - DIALOG OneSearch

File 5:Biosis Previews(R) 1969-2002/Sep W1
(c) 2002 BIOSIS

***File 5: Alert feature enhanced for multiple files, duplicates removal, customized scheduling. See HELP ALERT.**

File 35:Dissertation Abs Online 1861-2002/Aug
(c) 2002 ProQuest Info&Learning

File 65:Inside Conferences 1993-2002/Sep W2
(c) 2002 BLDSC all rts. reserv.

File 144:Pascal 1973-2002/Sep W2
(c) 2002 INIST/CNRS

File 155:MEDLINE(R) 1966-2002/Sep W2

***File 155: Alert feature enhanced for multiple files, duplicates removal, customized scheduling. See HELP ALERT.**

File 399:CA SEARCH(R) 1967-2002/UD=13711
(c) 2002 American Chemical Society

***File 399: Use is subject to the terms of your user/customer agreement. Alert feature enhanced for multiple files, etc. See HELP ALERT.**

4/9/1 (Item 1 from file: 5)

DIALOG(R)File 5:Biosis Previews(R)
(c) 2002 BIOSIS. All rts. reserv.

13655648 BIOSIS NO.: 200200284469

P2P -R protein localizes to the nucleolus of interphase cells and the periphery of chromosomes in mitotic cells which show maximum P2P -R immunoreactivity.

AUTHOR: Gao Sizhi; Witte Michael M; Scott Robert E(a)

AUTHOR ADDRESS: (a)University of Tennessee Health Science Center, 3 North Dunlap Street, Room N206, Memphis, TN, 38163**USA E-Mail: rescott@utmem.edu

JOURNAL: Journal of Cellular Physiology 191 (2):p145-154 May, 2002

MEDIUM: print

ISSN: 0021-9541

DOCUMENT TYPE: Article

RECORD TYPE: Abstract

LANGUAGE: English

ABSTRACT: P2P -R is a nuclear protein that can bind both p53 and Rb1. Its functions include roles in the control of RNA metabolism, apoptosis, and p53-dependent transcription. The expression of P2P -R also is repressed in G1 arrested terminally differentiated cells. The current studies therefore evaluated if P2P -R undergoes cell cycle-associated changes in its abundance and/or localization. Western blots show that relative to G0 quiescent cells, P2P -R protein levels are higher in

THIS PAGE BLANK (USPTO)

populations of G2/M cells prepared by the physiological parasynchronization technique of serum deprivation followed by serum stimulation. More striking is the > 10-fold enrichment of P2P -R protein in specimens of highly purified mitotic cells prepared by the mitotic shake-select technique, or by synchrony with the mitotic spindle disruption agents nocodazole or vinblastine. These changes in P2P -R protein occur without a concomitant change in P2P -R mRNA expression suggesting that P2P -R immunoreactivity increases during mitosis. Confocal microscopy next established the localization of P2P -R to nucleoli in interphase cells and at the periphery of chromosomes in mitotic cells that lack nucleoli. The high levels of P2P -R localized to the periphery of chromosomes in mitotic cells suggest that P2P -R shares characteristics with other nucleolar proteins that associate with the periphery of chromosomes during mitosis. These include: nucleolin, B23, Ki67, and fibrillarin.

4/9/2 (Item 2 from file: 5)
DIALOG(R)File 5: Biosis Previews(R)
(c) 2002 BIOSIS. All rts. reserv.

13653965 BIOSIS NO.: 200200282786

cDNA encoding P2P proteins and use of P2P cDNA derived antibodies and antisense reagents in determining the proliferative potential of normal, abnormal, and cancer cells in animals and humans.

AUTHOR: Scott Robert E(a)
AUTHOR ADDRESS: (a)Memphis, TN**USA
JOURNAL: Official Gazette of the United States Patent and Trademark Office
Patents 1257 (2):pNo Pagination Apr. 9, 2002
MEDIUM: e-file
PATENT NUMBER: US 6368790 PATENT DATE GRANTED: April 09, 2002 20020409
PATENT ASSIGNEE: University of Tennessee Research Corporation
PATENT COUNTRY: USA
ISSN: 0098-1133
DOCUMENT TYPE: Patent
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: Terminal differentiation is associated with repression in the expression of the P2P subset of hnRNP proteins. The present invention provides a P2P cDNA that encodes proteins with domains for hnRNP association and Rb1 binding. Probes to the P2P cDNA detect a single 8 kb mRNA in multiple murine tissues, in proliferating murine 3T3 cells but not in terminally differentiated 3T3T adipocytes. The interaction of P2P cDNA products and Rb1 may serve to modulate cell proliferation and/or other biological functions associated with tumor suppression by an RNA processing mechanism. Moreover, it was shown that P2P antisense oligonucleotides selectively repressed 30-40 kDa P2P expression.

4/9/3 (Item 3 from file: 5)
DIALOG(R)File 5: Biosis Previews(R)
(c) 2002 BIOSIS. All rts. reserv.

10834146 BIOSIS NO.: 199799455291

The proliferation potential protein-related (P2P -R) gene with domains encoding heterogeneous nuclear ribonucleoprotein association and Rb1 binding shows repressed expression during terminal differentiation.

AUTHOR: Witte Michael M; Scott Robert E
AUTHOR ADDRESS: Dep. Pathol., Univ. Tennessee Med. Cent., Memphis, TN 38163**USA
JOURNAL: Proceedings of the National Academy of Sciences of the United States of America 94 (4):p1212-1217 1997
ISSN: 0027-8424
RECORD TYPE: Abstract
LANGUAGE: English

ABSTRACT: Terminal differentiation is associated with repression in the expression of the proliferation potential proteins (P2P) subset of heterogeneous nuclear ribonucleoprotein (hnRNP) proteins. We report here the cloning and characterization of a 5173-bp P2P -related (P2P -R) cDNA that contains a 4214-bp open reading frame. Probes to this cDNA detect a single 8-kb mRNA in multiple murine tissues and in proliferating

THIS PAGE BLANK (USPTO)

3T3T cells, but not in terminally differentiated 3T3T adipocytes. Evidence that this cDNA can encode peptides with domains for hnRNP association was established by showing that such peptides are recognized by two monoclonal antibodies known to detect core hnRNP proteins, and by showing that the C130 monoclonal antibody, produced against a cDNA-derived fusion protein, also selectively detects native **P2P** hnRNP proteins. In addition, **P2P** -R cDNA-derived fusion proteins bind single-stranded nucleic acids, and a **P2P** -R cDNA-derived antisense oligonucleotide selectively represses **P2P** expression. Because terminal differentiation is associated with modulation in Rb1 function, we assayed if products of this cDNA might interact with Rb1. Evidence that the **P2P** -R cDNA encodes a protein domain that binds Rb1 was established using a glutathione S-transferase fusion protein to selectively precipitate Rb1 from cellular extracts. Data also show that this binding is reduced by competition with the adenovirus Eta protein, indicating that binding occurs through the "pocket" domain of Rb1. These results establish that the **P2P** -R cDNA encodes protein domains involved in both hnRNP association and Rb1 binding and complement recent reports that localize Rb1 to sites of **RNA** processing in the nucleus.

4/9/5 (Item 1 from file: 399)
DIALOG(R) File 399:CA SEARCH(R)
(c) 2002 American Chemical Society. All rts. reserv.

136261811 CA: 136(17)261811a PATENT
P2P protein, cDNA, antibodies and antisense oligonucleotides for determination of proliferative potential of normal, abnormal, and cancer cells in animals and humans
INVENTOR(AUTHOR): Scott, Robert E.
LOCATION: USA
ASSIGNEE: The University of Tennessee Research Corporation
PATENT: U.S. Pat. Appl. Publ. ; US 20020035080 A1 DATE: 20020321
APPLICATION: US 811045 (20010316) *US PV27568 (19960927) *US 801308 (19970218)
PAGES: 32 pp., Division of U. S. Ser. No. 801,308. CODEN: USXXCO
LANGUAGE: English CLASS: 514044000; A61K-048/00A; A61K-038/17B;
C07K-014/435B; C07H-021/02B; C07H-021/04B
SECTION:
CA215002 Immunochemistry
CA203XXX Biochemical Genetics
CA209XXX Biochemical Methods
CAS REGISTRY NUMBERS:
405051-29-8P amino acid sequence; P2P proliferation potential proteins, antibodies, nucleic acids and antisense oligonucleotides for diagnosis and treatment of proliferative diseases
405051-30-1P nucleotide sequence; P2P proliferation potential proteins, antibodies, nucleic acids and antisense oligonucleotides for diagnosis and treatment of proliferative diseases
405006-95-3 P2P proliferation potential proteins, antibodies, nucleic acids and antisense oligonucleotides for diagnosis and treatment of proliferative diseases
405051-74-3 unclaimed nucleotide sequence; p2P protein, cDNA, antibodies and antisense oligonucleotides for detn. of proliferative potential of normal, abnormal, and cancer cells in animals and humans

4/9/6 (Item 2 from file: 399)
DIALOG(R) File 399:CA SEARCH(R)
(c) 2002 American Chemical Society. All rts. reserv.

132304218 CA: 132(23)304218z JOURNAL
Gene-target recognition among members of the Myc superfamily and implications for oncogenesis
AUTHOR(S): O'Hagan, Ronan C.; Schreiber-Agus, Nicole; Chen, Ken; David, Gregory; Engelman, Jeffrey A.; Schwab, Richard; Alland, Leila; Thomson, Cole; Ronning, Donald R.; Sacchettini, James C.; Meltzer, Paul; DePinho, Ronald A.
LOCATION: Department of Adult Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA
JOURNAL: Nat. Genet. DATE: 2000 VOLUME: 24 NUMBER: 2 PAGES: 113-119
CODEN: NGENEC ISSN: 1061-4036 LANGUAGE: English PUBLISHER: Nature

THIS PAGE BLANK (USPTO)

CAS REGISTRY NUMBERS:

74812-49-0 gene Cull1 subunit Cullin 1; gene-target recognition among members of Myc superfamily and implications for oncogenesis - common targets for Myc and Myc(Bx11-basic region) chimeric protein
9024-60-6 gene ODC1; gene-target recognition among members of Myc superfamily and implications for oncogenesis - common targets for Myc and Myc(Bx11-basic region) chimeric protein
9013-05-2 gene P2R4; gene-target recognition among members of Myc superfamily and implications for oncogenesis - Myc-specific targets
9026-43-1 gene RING3; gene-target recognition among members of Myc superfamily and implications for oncogenesis - Myc(Bx11-basic region) chimeric protein-specific targets
120178-12-3 gene Tert; gene-target recognition among members of Myc superfamily and implications for oncogenesis - common targets for Myc and Myc(Bx11-basic region) chimeric protein
80449-01-0 gene TOP1; gene-target recognition among members of Myc superfamily and implications for oncogenesis - Myc-specific targets
74870-74-9 gene UMPS; gene-target recognition among members of Myc superfamily and implications for oncogenesis - Myc-specific targets
4/9/7 (Item 3 from file: 399)
DIALOG(R)File 399:CA SEARCH(R)
(c) 2002 American Chemical Society. All rts. reserv.

111229734 CA: 111(25)229734a JOURNAL

Loss of proliferative potential during terminal differentiation coincides with the decreased abundance of a subset of heterogeneous ribonuclear proteins

AUTHOR(S): Minoo, Parviz; Sullivan, William; Solomon, Larry R.; Martin, Terence E.; Toft, David O.; Scott, Robert E.

LOCATION: Dep. Biochem. Mol. Biol., Mayo Clin./Found., Rochester, MN, 55905, USA

JOURNAL: J. Cell Biol. DATE: 1989 VOLUME: 109 NUMBER: 5 PAGES: 1937-46 CODEN: JCLBA3 ISSN: 0021-9525 LANGUAGE: English

SECTION:

CA213006 Mammalian Biochemistry

IDENTIFIERS: ribonucleoprotein cell differentiation, proliferation potential protein cell differentiation

DESCRIPTORS:

Proteins, specific or class, P2P (proliferation potential protein)...

Ribonucleoproteins, heterogeneous nuclear RNA-contg....

of cell of human and lab. animal in terminal differentiation

Mesenchyme, stem cell...

ribonucleoproteins of, in terminal differentiation

Cell nucleus...

ribonucleoproteins of, of cells of human and lab. animal in terminal differentiation

Skin, keratinocyte, composition...

ribonucleoproteins of, of human in terminal differentiation

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 15:21:20 ; Search time 63.16 Seconds

(Without alignments)
2469.088 Million cell updates/sec

Title: US-09-811-045a-1
Perfect score: 7374
Sequence: 1 MAEVDPMNKGAMLTNTGKY.....WRSWLGWRAPSSRNPS 1404

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 4922 | 66.7 | 1278 | 22 | AAAM39259 |
| 2 | 3497 | 47.4 | 973 | 22 | AAAM41045 |
| 3 | 560.5 | 7.6 | 155 | 22 | AAAG3720 |
| 4 | 542 | 7.4 | 1231 | 22 | ABBS9738 |
| 5 | 531 | 7.2 | 146 | 22 | AAO02604 |
| 6 | 446 | 6.0 | 153 | 21 | AAAT0234 |
| 7 | 352 | 4.8 | 3910 | 14 | AAAR38470 |
| 8 | 352 | 4.8 | 3910 | 16 | AAAR66462 |
| 9 | 351 | 4.8 | 2768 | 22 | ABBB68397 |
| 10 | 344.5 | 4.7 | 3111 | 22 | ABBB60327 |
| 11 | 340 | 4.6 | 2703 | 22 | ABBB60074 |

| | | | | | | |
|----|-------|-----|------|----|-----------|--------------------|
| 12 | 340 | 4.6 | 3660 | 22 | ABG12812 | Novel human diago |
| 13 | 335.5 | 4.5 | 1026 | 22 | AAAM78925 | Human protein seq |
| 14 | 333.5 | 4.5 | 2519 | 22 | ABG16636 | Novel human diago |
| 15 | 332.5 | 4.5 | 1151 | 22 | ABBB1598 | Drosophila melanog |
| 16 | 331 | 4.5 | 2897 | 22 | ABBB5814 | Drosophila melanog |
| 17 | 331 | 4.5 | 2951 | 22 | ABBB60291 | Drosophila melanog |
| 18 | 330 | 4.5 | 1430 | 22 | ABBS8724 | Drosophila melanog |
| 19 | 329.5 | 4.5 | 617 | 22 | ABBS5445 | Peptide #2951 enco |
| 20 | 329.5 | 4.5 | 617 | 22 | ABBB20884 | Protein #2883 enco |
| 21 | 329.5 | 4.5 | 617 | 22 | AAAM56269 | Human brain expres |
| 22 | 329.5 | 4.5 | 617 | 22 | AAAM16458 | Peptide #2892 enco |
| 23 | 329.5 | 4.5 | 617 | 22 | AAAM04187 | Peptide #2869 enco |
| 24 | 328 | 4.4 | 1033 | 22 | AAAM79809 | Human protein seq |
| 25 | 328 | 4.4 | 2819 | 22 | AAAB35408 | Human OTCG27 gene |
| 26 | 326.5 | 4.4 | 1341 | 21 | AAAB56557 | Human Aclous L pro |
| 27 | 326.5 | 4.4 | 1342 | 22 | ABG09943 | Novel human diago |
| 28 | 324.5 | 4.4 | 2663 | 22 | AAAB98612 | Human tumour suppr |
| 29 | 324 | 4.4 | 2724 | 22 | ABG20119 | Novel human diago |
| 30 | 323.5 | 4.4 | 1333 | 22 | ABBB6754 | Drosophila melanog |
| 31 | 323 | 4.4 | 1524 | 22 | ABG04845 | Novel human diago |
| 32 | 321 | 4.4 | 3969 | 15 | AAAB52971 | Product of the cDN |
| 33 | 318 | 4.3 | 1163 | 22 | AAU28028 | Novel human secret |
| 34 | 318 | 4.3 | 5533 | 22 | ABBB5772 | Drosophila melanog |
| 35 | 318 | 4.3 | 5560 | 22 | ABBB7160 | Drosophila melanog |
| 36 | 315.5 | 4.3 | 3080 | 22 | ABBB4877 | Drosophila melanog |
| 37 | 309 | 4.2 | 3257 | 22 | ABBB7502 | Drosophila melanog |
| 38 | 309 | 4.2 | 6815 | 22 | ABBB6811 | Drosophila melanog |
| 39 | 306.5 | 4.2 | 935 | 21 | AAAB4231 | Human ORFX ORF195 |
| 40 | 306.5 | 4.2 | 962 | 22 | ABBB11976 | Human TRAP150 homo |
| 41 | 304.5 | 4.1 | 1183 | 22 | ABBB58769 | Drosophila melanog |
| 42 | 304 | 4.1 | 931 | 22 | ABBB1093 | Drosophila melanog |
| 43 | 303 | 4.1 | 724 | 21 | AAAG46505 | Arabidopsis thalia |
| 44 | 303 | 4.1 | 778 | 21 | AAAG46504 | Arabidopsis thalia |
| 45 | 301.5 | 4.1 | 1984 | 22 | ABBB1060 | Drosophila melanog |

ALIGNMENTS

| | |
|-----------|--|
| RESULT | 1 |
| AAAM39259 | standard; Protein: 1278 AA. |
| ID | AAAM39259 |
| XX | AAAM39259; |
| AC | 22-OCT-2001 (first entry) |
| DT | Human polypeptide seq ID NO 2404. |
| XX | |
| DE | Human: nontropic; immunosuppressant; cytostatic; gene therapy; cancer; |
| XX | peripheral nervous system; neuropathy; central nervous system; CNS; |
| KW | Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; |
| KW | amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; |
| KW | chemokine; thrombolytic; drug screening; arthritis; inflammation; |
| KW | leukemia. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | WO20015312-A1. |
| XX | |
| PD | 26-JUL-2001. |
| XX | |
| PR | 26-DEC-2000; 2000WO-US34263. |
| XX | |
| PR | 21-JAN-2000; 2000US-0488725. |
| XX | |
| PR | 25-APR-2000; 2000US-0552317. |
| XX | |
| PR | 09-JUL-2000; 2000US-0598042. |
| XX | |
| PR | 19-JUL-2000; 2000US-0620312. |
| XX | |
| PR | 03-AUG-2000; 2000US-0653450. |
| XX | |
| PR | 14-SEP-2000; 2000US-0662191. |
| XX | |
| PR | 19-OCT-2000; 2000US-0693036. |
| XX | |
| PR | 29-NOV-2000; 2000US-0727344. |
| XX | |

PA (HYSE-) HYSEO INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
DR WPI: 2001-44253/47.
DR N-PSDB: AA158415.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PS Example 4: SEQ ID NO 2404; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA158642-AA162213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 1278 AA:

Query Match 66.7%; Score 4922; DB 22; Length 1278;
Best Local Similarity 75.2%; Pred. No. 1.5e-286;

Matches 972; Conservative 81; Mismatches 148; Indels 92; Gaps 15;

QY 1 MMEVKDPNNKGAALNTWTGKAIPITDAEVAIGKKKKPPLEPESSSEEDDPIFAELL 60
DB 23 mmevkdpnnkgaaalntwtgkaipitdaeayaigkkkppllepepsseeddpiidell 82
QY 61 CLICKDINTDAVVIPCCGNSSCDECIKRTTLESKHTCPTCHONDVSPDALANKFLRQA 120
DB 83 clickdintdavyipccgnssycdecirtaallesdehtcpcqndvspdalankflrqg 142
QY 121 VNNFKETGTGTRLRKQOLPPFLFVPPRPRLSQRLQPRRSRITLQOQPVVRVRYVSP 180
DB 143 vnnfketygtkrlrkqlpppprpllqrlqlmrspisrvgdplmipvtssst 202
QY 181 CSDTKTAGSCSDSGTSLRPLPAPISISLTSSNQSILAPVSGNPSAPAPVDIATVTSISV 240
DB 203 -----hpapsissltngssilappvsgnpsapapvdiaatvtsisv 244
QY 241 HSEKSDGPRDSDNKLPLPAALITSEHSKGAASSIATITALEEKG---VPGTSPWNSIFWG 296
DB 245 hseksdgprrdsdnkllppaalitsehskgassiatitalmeekgyvplvgtpse---llg 300
QY 297 QSLTHQQLPTTGQPVAVINARBGGRPGWGHSHKGLVYSPPOQIRGERSCSRINRGR 356
DB 301 qslthqqlpttgqvavinarbggrpgwghshkglvlysppqqlrrgescyrinrgr 360
QY 357 HHSERQORTQSPPLPAPCPGVPPPPPLVPPPHLPLPPGVPPPOFSPQFSSOPTG 416
DB 361 hhsersqrqgspplpactpvgvpppplyppphlplppgvpppofspfpfp9gqatag 420
QY 417 YSVPPGFPAPANISTACFSPGVPAHNSNTMTTQAPLISREEFYREQ-----465
DB 421 ysvppgfpapanistacfspegvpaahnsntmttqaplisreefyreqrllkeekks 480
QY 466 -----NDKGRF-----SKPPYSGSSTSRSSYTDSS-SQGLAQHIAULT 501
DB 481 kldetfndfakelmeykklqkertrrsftrskapsygsysrseyltyskatsrstrsrsys 540

QY 502 LSPSAAHITLIDLHDPHP-----PEEAARSAMIVHMPDLMDIAHARSRPYRRYSR 557
DB 541 rfsfshsrsysrpppyrrrgskrnyrsrsh-----glynrsrsrpppyrrysr 595
QY 558 SPPEFRGQSPFKRNVPREKEKREYNRYREVPYDIKAYGSRVDPDRDFEKEERYREWE 617
DB 596 spqafrrgspnkrnvpqgetereyfnryrevpppydmkaygrsvdfdrpfekeyrewe 655
QY 618 RKRYREYKRYKGYAAGAPRPSANREDPSPEHLPLNINSPFTGRRRDYAAAGSHRW 677
DB 656 rkyreyekyygayaagaprpasarenentsperflplninspftgrredyggshrs 715
QY 678 RNLGNYPEKLTSTRDSHNAKNPKSKKESEENYPGDGKGNKHKHRRRNEKESESE 737
DB 716 rnlgnypelklsardghnqkndtkskeenaagdgknkhkhlrr---kgeesegf 772
QY 738 LNPELLFETSRKCSGSSGIDETKTDTLFLVPSRQDATPVYRDEPDASTIFKYSDDKRE 797
DB 773 lnpellfetsrksrepyveenktdslfvlpsridatpvrddepdaeslftksvsekdkre 832
QY 798 KDKPKVSPDKTKRKSQGSATAKKDNVLKPSKGPQEVYDGDREKSPSEPLKKAKSEATK 857
DB 833 rdkpkakgdktkrkdngsavskenlvkpkqpgkevderetspseppikakaeclpk 892
QY 858 IDSVKPSSSQKDEKVTGTPRKASHKSAKDTRRQSPRRRSKRYVPKTSQKSPVRR 917
DB 893 tdtksesssqkdekltgtrpkahskakehgetkpvkeekvkkdysk--dvksekltk 950
QY 918 RPRSLAKINYLAREKN-----EREKKR---KSYDKPFESSMKSIVEGEITYKPS 967
DB 951 eeka-kkpn-----eknkpldnkgekrkkteekgydkdfesssmksiklevelvkpsp 1004
QY 968 KRKMEGDVEKLEKTPREKNDIASSTTPAKKIKILNREGKIKGNENASTKREPSEKLESTS 1027
DB 1005 krkmpdtekmdtrpekdki-slsapakkilnregkikgnstenisntkpekselests 1063
QY 1028 SKIKQEKVKAKARKVAGSEBSSSTLVDTYSTSGSPYRKSEKTDKRYVIKTMEEY 1087
DB 1064 skvqgekvgkvrkvtglegessstlvdststsgspyrkseektldkrtvltkmeey 1123
QY 1088 NNDNTAPAEVYIMIOVPOSKMDKDFESEEDVYKTTQPOYSVGKSSIIKNVTTSPSAT 1147
DB 1124 nndntapadvylimigvpskwkddfeseedvktqplissvgkpsavylknvstlkspsnl 1183
QY 1148 AKYTEKESDPEKLOKLPKASHELMQHELRSKGSASSKGRKAPREVSSEKNDPDR 1207
DB 1184 vkypekesepsekiqkftkdvshelilqnevksksnassekgyktkardysvlekenpekr 1243
QY 1208 KSGAQPDKESTVDRLSEOGHFKTLQSQSKETRT 1240
DB 1244 kstgpekesnldrlnegnfkslsgskeart 1276
RESULT 2
AA041045
ID AA041045 standard; Protein; 973 AA.
XX
XX AA041045;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 5976.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX


```

Db 531 aslygvaakvsgpiddpleafnrimkekerkkvdrfrssdrhsrpsd---rqrhrfk 587
QY 679 NLGNYP--EKLSTRDSHNAKD-NPKSKEKESENVPGDGKGNKKRH-RKRREKEGSES 734
Db 588 s-----pmekdnrd--nlkdkprsrerkrh-----syerhlrprsrqndgsks 635
QY 735 ESFLNPELETSRCKRGSSGIDETKTDLFLVLPFRDDATPYRDBPMOAEISTFHSVSKD 794
Db 636 pg-----gtrkr--sg-----hrtsasp----- 651
QY 795 KREKDKPRKVSDDKTRKSDGSAATKADNVLPKSG--POEKVDGDRKSPRSEPLKAK 852
Db 652 -----kpgyksdydkynkpsapktaevepppgfeqlqtd---edgynkptisea 703
QY 853 EEARLDSVSPSSSQDEKVTGTPRKAH-----SKSAKTRRDSQPTKRSKATVTK-T 906
Db 704 sqsgksgskkrngenrtee-----aprkhrsrstskpkrpdsnyrstltpakkltltpkmt 759
QY 907 SSQSKQPVTRPRRSLRKI--NYLIAREK-----NERKRRKSVYDKDESSSMKISK 956
Db 760 aeqrlr-----reespktepsknddylltakarlnasqpvlnrde-metnvgkenkaks-plsk 815
QY 957 VEGTEIVKPSPRKMEGVDEKLETPKDKIASSTTPAKKIKLNRGTGKIGMAENASTT 1016
Db 816 -----drkkkkdkdkkaer-----knkkdkkrakkekqdrqkks 851
QY 1017 KEPEKELETSISKQEKVKAKRKVAGSGSSSTLVDTYSTSGSPYRKSEKTDVT 1076
Db 852 vnrqd-----sdlnasslmessnykvlspraqpsl---elnaaqslpshnatenvp 901
QY 1077 KRTYITMEEYNNNDTAPAEVITIMIQVPOSKMDKDFESEEDVKTQPIQSGVKPSSI 1136
Db 902 kshsilllvgaasdnllprsklseansvnlskweidenllgled-sakkaagaaddpsel 960
QY 1137 IKNVTTTPSATAKTEKSEDEPEKLOKLPRKASHELMOHELRRSSGASSEKGRADREH 1196
Db 961 tsdvlfkaenalfakalnairtmefqyl-----lnskdnksdrsv 1000
QY 1197 SGSEKD---NDPKRRSGAQPRKESTVDRL-----SEQHFKTLSSOSKETTSKHESVR 1248
Db 1001 vrsqkdrsssprtns-----strvkdltgcklndrstrsdkskgyrraarssdddnr 1055
QY 1249 GSSNKDTPGGRKRVYDYSRDYSS-SKRROREGELARKKSDPPGKESLSGOKSLRPER 1307
Db 1056 gsrshgsrktndr-----strdaapekrgrer---sykrirsp-----eddkllrrqn 1099
QY 1308 DLPRKGAESK---KSNSSPPRDKKPHDHKAPYETK-----RPECETKP--- 1347
Db 1100 ---kegseskhgkhdqnsddsdtrraakntksdgsrsvsvtlavappkpcrpnpfkff 1156
QY 1348 VDRKNSGRK-----EKHAAEARNGKE 1368
Db 1157 vdtssssslvvkynltlqkegassdngme 1185

```

```

RESULT 5
AA002604
ID AA002604 standard; Protein: 146 AA.
XX
AC AA002604;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 16496.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN MO200164835-A2.

```

```

XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI: 2001-514838/56.
XX N-PSDB: AAI82535.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune
XX PT disorders -
XX
XX Claim 20; SEQ ID NO 16496; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 146 AA;

```

```

Query Match 7.28; Score 531; DB 22; Length 146;
Best Local Similarity 71.5%; Pred. No. 1.2e-24;
Matches 103; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
QY 1041 RKVGSSESSSTLVDTYSTSGSPYRKSEKTDTRKTVIKTMEYNNNDTAPAEVIT 1100
Db 3 rkvgtegsstvlqyltpsstgmpvrksdqctkvtlktmedynndtapaedvii 62
QY 1101 MIOVPOSKMDKDFESEEDVKTQPIQSGVKPSSIKNVTTTPSATAKTEKSEDEPEK 1160
Db 63 mlyqpslwdqddesedqddksqslssvgkpatlvknvtrtkpsalvkkypekesepfak 122
QY 1161 LQKLPRKASHELMOHELRRSSGSA 1184
Db 123 llkfmdvshvlphevksxsyxa 146

```

```

RESULT 6
AAI70234
ID AAI70234 standard; Protein: 153 AA.
XX
AC AAI70234;
XX
DT 06-JUN-2000 (first entry)
XX
DE Human RNA-associated protein-15 (RNAAP-15).
XX
KW RNA-associated protein; RNAAP; human; clone 1879135; cytosolic;
KW immunosuppressive; antiinflammatory; keratolytic; neuroprotective;
KW antiarteriosclerotic; hepatotropic; antiproliferative; virucide; anti-HIV;
KW antiallergic; antihemmatic; antiarthritic; ophthalmological; autoimmune;
KW antimicrobial; cell proliferative disorder; inflammation; cirrhosis;
KW actinic keratosis; bursitis; arteriosclerosis; artherosclerosis;
KW hepatitis; myelofibrosis; primary thrombocytopenia; psoriasis; cancer;

```

KW mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;
 XX allergy; rheumatoid arthritis; parasitic infection.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FT Modified-site 5
 FT Domain /note= "Potential phosphorylation site"
 FT 26..43
 FT /label= zinc_finger
 FT Modified-site 43
 FT /note= "Potential phosphorylation site"
 FT Modified-site 59
 FT /note= "Potential phosphorylation site"
 FT Modified-site 82
 FT /note= "Potential phosphorylation site"
 FT Modified-site 112
 FT /note= "Potential phosphorylation site"
 FT Modified-site 113
 FT /note= "Potential phosphorylation site"
 FT Modified-site 114
 FT /note= "Potential phosphorylation site"
 FT Modified-site 145
 FT /note= "Potential phosphorylation site"
 FT /note= "Potential phosphorylation site"
 PN WO200011171-A2.
 XX 02-MAR-2000.
 XX 20-AUG-1999; 99WO-US19361.
 XX 21-AUG-1998; 98US-0097550.
 XX 12-JAN-1999; 99US-0115639.
 XX (INCY-) INCYTE PHARM INC.
 XX Hillman JL, Yue H, Tang YF, Corley NC, Guegler KJ, Gorgone GA;
 PI Paterson C, Baughn MR, Lal P, Bandman O, Reddy R, Azimzai Y;
 PI Shih LT, Yang J, Lu DM;
 XX WPI: 2000-237651/20.
 DR N-PSDB: AA51264.
 XX Human RNA-associated proteins useful in diagnosing, treating and
 PT preventing cell proliferative, autoimmune, inflammatory and infectious
 PT disorders -
 XX Claim 1: Page 94-95; 123pp; English.
 XX The present amino acid sequence is the human RNA-associated protein-15
 CC (RNAp-15), identified in Incyte clone 1879135, derived from LEUKNO703
 CC library. It is expressed in reproductive, dermatologic, gastrointestinal,
 CC musculoskeletal and haematopoietic/immune tissues. It has cytosolic,
 CC immunosuppressive, antiinflammatory, antiarteriosclerotic, hepatotropic,
 CC keratolytic, neuroprotective, antipsoriatic, anti-HIV, anti-allergic,
 CC antirheumatic, virucide, antiarthritic, ophthalmological and antimicrobial
 CC activity. RNAp antibodies are useful for diagnosis of diseases
 CC associated with altered expression or activity of RNAp. It is used to
 CC treat cell proliferative, autoimmune, inflammatory and infectious
 CC disorders, like actinic keratosis, bursitis, arteriosclerosis,
 CC artherosclerosis, cirrhosis, hepatitis, myelofibrosis, mixed connective
 CC tissue disease (MCTD), psoriasis, primary thrombocytopenia and cancer,
 CC HIV, allergies, rheumatoid arthritis, uveitis, Crohn's disease, and
 CC bacterial, viral and parasitic infections.
 XX Sequence 153 AA:
 SO

Query Match 6.0%; Score 446; DB 21; Length 153;
 Best Local Similarity 97.6%; Pred. No. 1.7e-19;
 Matches 83; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MMEVKDPNMKGAMLTNTGXYAIFTIDAENVAIGKKKPPPLPEPSSSSSEDDPIPAELL 60

DB 66 mmevkdpnmkgamltntgxyaiftidaenvaigkkkppplpepssssseddpdpell 125
 QY 61 CLICKDITMDAVYIRPCGNSCDEC 85
 DB 126 clickdmtavvipcgnsycdec 150
 RESULT 7
 AAR38470
 ID AAR38470 standard; Protein: 3910 AA.
 XX AAR38470;
 AC AAR38470;
 XX 08-NOV-1993 (first entry)
 DT
 XX ALL-1 protein.
 DE
 XX Acute lymphoblastic leukemia gene; ALL-1; chromosome 11; treatment;
 KW translocation breakpoint mapping; chromosomal abnormality; diagnosis;
 KW human; acute lymphocytic; myelomonocytic; monocytic; myelogenous;
 KW leukemia; Drosophila; trithorax; homology region; zinc finger domain;
 KW cysteine-rich.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH Region 1021..1221
 FT Region /note= "Region of homology to Drosophila trithorax"
 FT Region 1462..1570
 FT /note= "Region of homology to Drosophila trithorax"
 FT Region 3348..3562
 FT /note= "Region of homology to Drosophila trithorax"
 XX MO9312136-A.
 XX 24-JUN-1993.
 XX 09-DEC-1992; 92WO-US10930.
 XX 11-DEC-1991; 91US-0805093.
 XX 27-MAY-1992; 92US-088839.
 XX 30-OCT-1992; 92US-0971094.
 XX (UYJE-) UNIV JEFFERSON THOMAS.
 XX Canaan E, Croce CM;
 PI WPI: 1993-214090/26.
 DR N-PSDB: AAQ43526.
 DR Detection and treatment of acute leukemia(s) - using prods.
 PT derived from oligo:nucleotide sequences within the ALL-1 gene of
 PT chromosome 11
 XX Disclosure; Page 29-50; 90pp; English.
 PS This sequence is encoded by the acute lymphoblastic leukemia (ALL-1)
 CC gene of chromosome 11. The ALL-1 gene was isolated by translocation
 CC breakpoint mapping. Fragments of the ALL-1 cDNA may be used to
 CC identify chromosomal abnormalities within the ALL-1 gene. These
 CC fragments may be used in the treatment and diagnosis of human
 CC leukemias such as acute lymphocytic, myelomonocytic, monocytic and
 CC myelogenous leukemia. ALL-1 protein shows three regions of homology
 CC to the Drosophila trithorax protein. These regions show 64%, 66% and
 CC 82% similarity respectively, to the Drosophila gene. The third region
 CC of homology constitutes the extreme C-terminus of the two proteins,
 CC both proteins end in an identical sequence. The first homology region
 CC is cysteine-rich and contains sequence motifs analogous to four zinc
 CC finger domains (3-6) within the trithorax gene. The second region of
 CC homology is also cysteine-rich and corresponds to zinc fingers 7 and 8
 CC of the Drosophila gene. The multiple conserved cysteines and
 CC histidines at the 3' end of the motifs allow two or three arrangements

| | | | |
|----|------|---|------|
| Db | 2308 | v--spvfdvrepsekeprav-----seydgeescepryhndetsepts----- | 2349 |
| Qy | 1120 | DVKTTQPIQSVGKPS-----SLIKNVTTKPSATAKYTERESEQPEKLQKLPRKEASHELM | 1173 |
| Db | 2350 | daklkprtsaparpseepateaeiypetaapelekevepateqepelakepate--- | 2406 |
| Qy | 1174 | OHELRSSKGSASSEKAKRAKRREHSGSEKONPPKRRSGAOPKESIVYDLSQGHFKTISQ | 1233 |
| Db | 2407 | qpel-----eketpекateqepelеke--pекateq-----pe | 2437 |
| Qy | 1234 | SSKET--RTSEKHEVYKGSNKDPFTGPRDKYDVOSRDVSSSKRDEKGELARRKDSPPR | 1291 |
| Db | 2438 | leketvkatateqesv-----dektlpepyvbkpeldstedeesvsee--esaadkd---k | 2489 |
| Qy | 1292 | GKESLSQKSKLLEERDLPKKGA--SKKSNSSPPRDKP--HDHKAPEYTKRRCETKRPVD | 1349 |
| Db | 2490 | nket--eedtkkhneepvavvseipqpeaeavpttghpelfhlaasstltpra-----vd | 2543 |
| Qy | 1350 | KNSGKEREXH 1359 | |
| Db | 2544 | drwygeedan 2553 | |

RESULT 10
 ABB60327 standard; Protein; 3111 AA.
 ID ABB60327 standard; Protein; 3111 AA.
 XX
 AC ABB60327;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 7773.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 XX
 PD 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US09231.
 XX
 PF 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 PR
 XX (PEKE) PE CORP NY.
 PA
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 PI
 DR WPI: 2001-656860/75.
 DR
 N-PSDB; ABL04430.
 XX
 XX
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX
 PS Disclosure; SEQ ID NO 7773; 21pp + Sequence Listing; English.
 PS
 XX
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX

sq Sequence 3111 AA;

| | | | | |
|-----------------------|-------------------|--------------------|-------------|-------------|
| Query Match | 4.7% | Score 344.5; | DB 22; | Length 311; |
| Best Local Similarity | 19.1%; | Pred. No. 1.3e-11; | | |
| Matches 264; | Conservative 412; | Mismatches 172; | Indels 535; | Gaps 55 |

| | | | | |
|----|------|---|---|-------------------|
| OY | 442 | TAHNSWPTTQAPLRLSNEEYREQDNKGR- | -----SKPRYSSSVSRSSYDSQGLAQ | 495 |
| Dd | 669 | tslerqprksvapshaprfmflmgtdsgdckstkeqprltygdq | aedstates--ghye | 725 |
| OY | 496 | HIHALTSPSAHTLDLIDH- | -----PHEPEAEARSAMI- | 530 |
| Dd | 726 | fdrelhlvskmepklpfdpbeftgmtsvspritprekseaeedl | lderckpfdvsddgv | 785 |
| OY | 531 | -----VAMPRLMIAIA- | -----RMSPPYRRRRSRSPRPF- | 562 |
| Dd | 786 | ngyfervnsverprlcltyslneeedtnaileeekdvghnsqdr | reynslpptypren | 845 |
| OY | 563 | -----RGOSPTKR--NVPREKEREFNRYRV- | -----PPYDIKAYGNSV | 602 |
| Dd | 846 | ddgyvggkqvptidndhrgenk | -----dykllsmtleentlvuypprptastlsnar | 900 |
| OY | 603 | DFRDPFEK- | -----ERYMEKTYEMTEKYTGAGVADPRPSANEDSPERL | 651 |
| Dd | 901 | krtprrtklltrsltleerygalertl | -----sqdqpssdq | 939 |
| OY | 652 | LPLMIRNSPFRGAREY- | -----AAGSHRNMLGNYPERL-----STRD-- | SHNA 696 |
| Dd | 940 | -----akyipstaaleefnclqklsaeqprklsameeyrpkser | lpslstdlsfrns | 995 |
| OY | 697 | KDNPKRSKESENVPGD- | -----GKGNKKHKRRRNE-EKGESESLNPLELTSR | 747 |
| Dd | 996 | ltkqmsessssckprldlkdedrpsgsskngqdxektskl | lnkseepentkettgetea | 1055 |
| OY | 748 | KCRSSSGIDERTD--LFLVLP- | -----RDA | 772 |
| Dd | 1056 | sdsndsklgeketepqrklkprltaeldfinalerkmvsgksspaknkkeppdeeks | ----- | 1135 |
| OY | 773 | TPVDEPDADSTIFKS- | -----VSDDKRREK- | -----KPKVKSDT 808 |
| Dd | 1116 | tkeepeesekanehtcsqgtrptlakkdsdsqkksetekemqsp | lknqdvkvtkpsks | 1175 |
| OY | 809 | KR--KSDGSATAKDNVLAKSKGQOEKVDGRE- | -----kspRSPPYLKAKKEA- | 855 |
| Dd | 1176 | eemleketssnpkeds--heseaatlkkvgegnrlssekgdnh | lkeksseaprgkagketae | 1234 |
| OY | 856 | TKDISVPSS--SSOKDE--KVTGT | -----PRAHKSXAKDTRQOSQ- | 895 |
| Dd | 1235 | tknanvksdkkqgdqkneaktsvsqgtesdlkpsksenstakdaeqekt | trpkpsptee | 1294 |
| OY | 896 | -----TRSKRTVPKTSQK--SQPVTRRRP- | -----SLRKINYLIA | 930 |
| Dd | 1295 | lekrfnalekqmtlnletlekepqdtkpraktksqstaevtktksmksf | adklkevnale | 1354 |
| OY | 931 | REKN--EREKRKSIV- | -----DKOFE | 948 |
| Dd | 1355 | keqsrvevevnaekkrkvnveaapknkeqdsqgqeesqhkqknq | rtasrepstdlekye | 1414 |
| OY | 949 | SSSKMKIS- | -----VEGTIVPSFKRK-----MEGD- | 974 |
| Dd | 1415 | tlkrmskshqfsetvealerlqgevlseaveekppstdlesr | fialnlgdkknes | 1474 |
| OY | 975 | ----- | -----VEKLE-----KPEKDK | 986 |
| Dd | 1475 | kmdekhvdaaleahpsppppppkperpvlaeprlhqgqal | leelqskmrgsgpsen | 1534 |
| OY | 987 | I-ASTTPARKIK- | ----- | 998 |
| Dd | 1535 | lkpeinpqrtrqkllqyrptmgdetsaeapantayyaa | nhbqwgqmvrtfsldprad | 1594 |
| OY | 999 | -----LNNETGKKI-----GNAENASTTKEPSEKLESTSSKIKOEKVKGAKARRVAG | ----- | 1045 |

Db 1595 lenrlqflerlykkykqrcasdevasryklppedqstargarkgeagqlegrvia 1654
QY 1046 SEGSSS-----TLVDTSTSSGSGSPVRSSEKTDTKRIVITMEYNNNDNAPAD 1097
Db 1655 lekqlsensklleamrerhrsaddgsprlsetld-----atgke 1697
QY 1098 VIMIQVQSKMDKDFESEDVKTQPIQSVGKPSST---IKNVTTPSATAKYTEKE 1154
Db 1698 lv-----ytqnlgeeevdah-----xpinishikmmvknks-----e 1732
QY 1155 SEQP-----EKLQKLPKASHHELMQHELRSKGSASSEKRAKDRHSGSEKNDPDK 1206
Db 1733 skpgyesktptedlrrtleqleqll--eerakngslipenevleekpekileedsck 1790
QY 1207 KKSQA--QPDKSTYDRLSEGHFRTLSSSKETRTSEKHESVR-----GSSKDDT 1256
Db 1791 geknchnqnykgydevekteipadrklepasakektlenvekaqtrakvvdtekskddn 1850
QY 1257 PGRDKKVDSDRYSSSKRDERGELARRKRDSPPRGK---ESLSGQSK---LREERDL 1309
Db 1851 avtdeksvqdgqnyvvdkka--drklldkdkspaagksedtkqtsqkksedlkqasea 1908
QY 1310 PKKGAESKKSNSPPRD---KKPHDHKAPYETKRPCCEETKRPVKNKSGKERKHAARNG 1366
Db 1909 pkagasketstrgkpscklekptkesylvketfpkkenlesekpkksenaatkietqks 1968
QY 1367 KES 1369
Db 1969 Ket 1971

RESULT 11
ABB60074
ID ABB60074 standard; Protein; 2703 AA.
AC ABB60074;
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 7014.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
OS Drosophila melanogaster.
FN WO200171042-A2.
PD 27-SEP-2001.
PF 23-MAR-2001; 2001WO-US09231.
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
DR WPI; 2001-656660/75.
DR N-PSDB; ABL04177.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PS Disclosure; SEQ ID NO 7014; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2703 AA;
Query Match 4.6%; Score 340; DB 22; Length 2703;
Best Local Similarity 20.0%; Pred. No. 2e-11;
Matches 277; Conservative 178; Mismatches 534; Indels 398; Gaps 63;
QY 95 KHTCPTCHQNDVSP---DALINKFLRQAVNNFKMETGTRKLRLQ--LPPLFLVPPPR 149
Db 723 qhqspr-ihqgqgsphqhqlhpsynsgqgfdpfaillgdnqpaagqqlmshvssypva 783
QY 150 PLSORNLQ-----PSSRSPTL-----RODPVYFRYVSPSCSDRTAGSCSDS 193
Db 784 qmqqgqhqgtqgqsgqgqgqgvsqgqgqgqgqpsrphhisslmet---ansnds 840
QY 194 GTLSRLPAPSISSLTSSNOSLIAPVSGNPSAPAPVDTATVSISSHSEKSDGPFRRS- 252
Db 841 stl-----avaandsnssnstnslnvngpvtshdngsqssinsnnqqlimndt 894
QY 253 -----DNKILP--AAALTSEHSKASSIATLTALMEENGVCSTFW 290
Db 895 nshgnsesagqagdvglfdnsmnsnaavaavasaasaa--alldnsngsnaef 952
QY 291 -----NSIFVGOSLHGLIPTGTPVRIN--AARPGGR 322
Db 953 ekngseeglyvvegyianetaldpdensvsktdtltstlept--lgddlqekptlaem 1009
QY 323 PGWEHSNKLIG---YLVSPPQQLRR-GERSCYRSINRGHNSRSQRTQSPSLPATPCFV 377
Db 1010 eqtlqigsldisksllleeklveesge-----dksqapgyvarepdr-ll 1054
QY 378 PYPPPLPYPPPHTLPLPPGVPPPOFSQFPSSQPTTACYSVP--PG-FPPAP----- 428
Db 1055 pvtqppqgqmpilgmhmblapgyeaqgqgqjmlmp--mmppyygagqmpypymllhq 1110
QY 429 ANISTRA-----CFSPVPTAHSNTMPTQAPLLSREEFYREQNDKREKFPVSGSS 480
Db 1111 getlaaqqqigelycmpprhelhqdklmrmgerl-----nl 1147
QY 481 YSRSSYTDSSQGLAQHIALTLSPSAARTLDLHDHPPEAEARSAMIVHMPDLMDIA 540
Db 1148 ltchevndqcaqgpcqllfgnvprmyg-----ppgnppllnqmv----- 1186
QY 541 HARSRPPTRRYRSRSRSPFERGOSPTKRNVPREKEREFTNRYREVPPTDIFAYVGR 600
Db 1187 ----espyvstltgry-----rgksaankprkprakke-----kayvqg 1221
QY 601 SVDFRDPFEKERYREWERKRYREWEKYKYGVAVGAOPRSANREDPSPRLPLNIRNP 660
Db 1222 qgdldm-----isgnvanga-----anaavslptqlp--vsecd 1254
QY 661 FTGRGRREDYAAQOSHRNRNLGNYPEKLSTRDSHNAKNDPKSKEKESEVPGDGKGNKIK 720
Db 1255 vtcgaqdtavgmleysegmd-----lsqdhhsaneidst-----dgsqkx-k 1298
QY 721 KHKRRNEKEGSESEFLNPELLETTRKCRGSSGIDETDTDLFLVPSRDDATPVVDEDM 780
Db 1299 kprkprtkp-----dpnk-----ppr-dtlpaxkpkp 1324
QY 781 DA-----ESITFKSVSDKDKREKDKPVKSDKTKRKSDGASATAKDNVLPKSGPOEKVYG 836
Db 1325 dnpndpstetpaavkkragskrtkygedgaetgegevev--edn--kplpkagsadg 1380
QY 837 DRKSPRSRPPLKKAKEAKTKIDSVKp-----SSSSQK-----DEKVTGTRR 878
Db 1381 eaett-----sgtlyvdgespdyddipvsklprgsnedekeagaagdetvdsvpd 1429

QY 879 KAHKSAKDTFRQSQPTFRKSKRTVPTKSSQKSPVTRRRRLKINYLIAKENEK 938
DB 1430 sagpastrprdrckrtatarr---nanseegsarknr-----gsIsakalkkrrn 1479
QY 939 RKKGV-RKDFSSSMKISKVGTETLVKPSPKRMKEGVEKLEPTEPKOKIKMS----- 990
DB 1480 rgrlvpesdgedatll-----drtppepppsdmsdnkrrsrrtqtkyldtvmlyfsd 1534
QY 991 -----TTPAKKIKLNRGTGKKIGNMENASTKEPSEKLESTSSKIKOEKVKGRKARKV 1043
DB 1535 densllvaspykkdkd-----kpsanaasnagsd-vekte-----pqsgaegdaagev 1581
QY 1044 AGSGSSSTLVDTY---STSSGSPVRAKSEKTDTKRTVTKTMEENNDMTAAEDYIT 1100
DB 1582 -geeksnlpidesqleassetsavaekergjstdaanaasakpnyyIntgdedsmv 1640
QY 1101 MIOVQSKMDKDFESEBEDVKTTPIQSVGKPSIIKKNVTKRSANAKY--TEKESQPE 1159
DB 1641 q1-vlammgkrrtllldkpkakepkyqdeek-seldeatldkpegdekltgees---- 1694
QY 1160 KLOKLPKEASHLMOHELRRSSKSGSASSEKGRAKDREHSGSEKDPDRK-----SGAQP 1214
DB 1695 -----kkdlitseeetklessamevdskeseepd-dsktsdeahkdkdkmevdeygsd 1747
QY 1215 KESTVDRLSE---QGHFKTISQSKETRTSEKHESVRGSSNKDFTPGRDKKVDYDSRDY 1270
DB 1748 keskpgeqsetrvkteenskaledkssvtladh-----akepetvlekmewdekand 1801
QY 1271 SSSRRRBERGELARRKDSPP-----RGKESL--SGOKSKIREERDLPRKKAESKKSNS 1322
DB 1802 gsaavsksegsdektcdsdnpaeatekkeslelegekervkegeevkxkxend---te 1858
QY 1323 PPRDKP 1329
DB 1859 admenkp 1865
RESULT 12
ABG12812
ID ABG12812 standard; Protein; 3660 AA.
AC ABG12812;
DT 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #12803.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
OS
XX WO200175067-A2.
PN 11-OCT-2001.
PD
XX 30-MAR-2001; 2001MO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR N-PSDB; AAS76999.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX
PS Claim 20. SEQ ID No 43171. 103pp. English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3660 AA.
Query Match 4.6%; Score 340; DB 22; Length 3660;
Best Local Similarity 20.1%; Pred. No. 2, 9e-11;
Matches 283; Conservative 153; Mismatches 463; Indels 512; Gaps 60;
QY 111 LIANKFLRQAVNNKNETGYTKRLKQLPP-----LFLVPPRPISQRLQ 157
DB 59 llmsnyrgq-----kkgqakklekaeqgkrkvtkyknlrqfllmpvsaaisrlk 111
QY 158 PRKSPILRQODPVY-----FRYVSPCTSDPKTKAGSCSD-----SGTSLRLPAP 202
DB 112 tprfiededydpiklarlestpnstfs-apsccsgksksaaqhsqmsdsstsssp 170
QY 203 SLSLTSNOS-----LAPVSGNPSSAPAPVPDITATVSISVSEKSDGPFRRSDKLL 257
DB 171 svdstsdsqaseelqvipeestdpevhp-plp-----lsgspenendrrrrys---- 220
QY 258 PAAALTSEHSGKASSIAITALMEKGVGTPSMNSIFVGSLHGLIPTTGPVRIANAR 317
DB 221 -----versfsgrt----- 230
QY 318 PGGRPRGHEHNSKRGIVLPPQOIRRGRCYRISNGRHHSEKORTQSPSLPATPCFV 377
DB 231 -----lklstlqsapqg-----qlsssp----- 249
QY 378 PVPPPLYPVPP-----HTLPLPGVPP--POFSPQFSSQPTAGVS--VPPG 423
DB 250 --ppplltppplqpassisdht--pwlmptlplaspflpaspstamgkkrksllrept 304
QY 424 F-----PPAPANISTAGFS-PGV--PTAHSNTPMTQAPLLSREEFYREONDKGRES 472
DB 305 frwtslksrsepqyfesakyakeglirxpifdftrppltp-----edvgfas 353
QY 473 KFPYSSGSSYRSSTYDSSQGLAQHIIHALTILSPSAHRLDLHLHPRHPEAEARSAMIVH 532
DB 354 gfsasglaaasarlfsplhsgrlfomh--krs-----llrprtfpsahsrllfesvt 404
QY 533 MPDLMDIAHARSRSPRYRRRSKRSRSPPEFGOSPTRKNVPRKEKEREVNRYREVPVY 592
DB 405 lpsnrtsaaglessgvsnrkkirkvfpisrepsps--hsmrttsqslsselspltpss 462
QY 593 DIRAYYGRSVD-----FRDPEKERYREMERKYREMYKYGVVGAQPRP--- 639
DB 463 svseslslsvslatsalnpltfps-----hsltgsgeaeaknqrprkq 507
QY 640 -SANREDF-----SPERLLPLMIRNSPFTGRGRREDVAAQOSHRNRLGNGNYEKLS-TRDSH 694


```

Db      508 tsapaeffssspplpfwftpgsqtergrnkda-----peelskdrdad 553
QY      695 NAKNPPSKKEESENWPDGKGKHKHKRRRNEKEGESESLNP-----ELL 743
Db      554 ksvekdksrerdrer-----ekenkresrkekrkgselqssallypvgvskvkgadv 609
QY      744 ETSKCKGSSGIDETKTDTLFLVLPSSDDATPVREPMASITFKSVSDKMKREKDKV 803
Db      610 atssasakaklqkss-----shdsqt-----dltsvli----- 638
QY      804 KSDTKRRKSDSATKADKNVL-KPSKGPOEKVDGDRKSPSEPLKAKE---EATKID 859
Db      639 -----gdtavatklikkgignlaktnd--lgr-tapslekleklclstpsss 685
QY      860 SVKSSSS-----OKDEKVTGTPPKHSAKSAKDTRRSOPTRRSKRTVEKTS 907
Db      686 tvkhstssigsmiaqadklpmtdkrvasllkkakqclcklkskqlqtdg----pkaq 740
QY      908 SOKSOPVAT--RRPRSLRKINYLAREKNEREKRRKSVDKOFESSSMKISKVEGTETVKP 965
Db      741 gqesdssetsvrgr-----lkh-vcrraavalgikravfpd----- 776
QY      966 SPKRMEGDVKELEBTP--EKDKIASSTTPPAKKIKLNRETCKKIGNAENASTTEPSEKL 1023
Db      777 -----dmpltsalpwereekllss-----mgndksslagsedaaplakpi 820
QY      1024 ES-TSSKTKOKK--VKGAKAKRVAGSEB-----SSSTLVDTSTSSGSGPVKRSSEKTD 1075
Db      821 kvprtnkagpepyvkkgrsrrcgcpcqcyredacqvcvctndlphkpfgrnlkkgcck-- 878
QY      1076 TKRTVITKMEYNNMDNPAADVILMIQVPSKWDKODFESEEDPVKTQPIQSVGKPS 1135
Db      879 -----mrkcnlqwmppskaylqkqakavkkkksktsckd-----skss 920
QY      1136 IIKNVTTPSATATKYTEKESDPEKLOKLPKASHELMQHELRSKSGSASSEKGRANDRE 1195
Db      921 vvkhnv-----dsqgkp----- 932
QY      1196 HSGSEKNDPDKRKSGAQDASTVDRLSEOGHFTLSOSSKETRTSEKHESVKGSSNKF 1255
Db      933 -tparedpakkssepprkpkvyeeksegnvasapgesqatpaarksskysqpal 991
QY      1256 -----TPGRDKKVDYDSRDYSSSKRRDERGELARRKDSPPRGESLSGQKSKLREERD 1308
Db      992 vlpqppttpprk-----evpkltpsepkkkqp-----p 1021
QY      1309 LPKKGAESKKSNSPPRDKKPHDHKAPYETKRPEETKRPVDK-----NS 1352
Db      1022 ppsqpsqskqkqvaprsipvkqk-pkekekp-----ppvnkgenagtlmlstlsngns 1076
QY      1353 GKER-----EKHAABARNGKSSSG 1371
Db      1077 skqktpadqyvhrlrvdfkedceaanvwmegm 1107

RESULT 13
AAAM78825
ID      AAAM78825 standard; Protein; 1026 AA.
XX
XX      AAAM78825;
AC
XX
XX      06-NOV-2001 (first entry)
DE
XX      Human protein SEQ ID NO 1487.
KW      Human; cytoline; cell proliferation; cell differentiation; gene therapy;
KW      vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW      tissue growth factor; immunomodulatory; cancer; leukemia;
KW      nervous system disorder; arthritis; inflammation.
XX
XX      Homo sapiens.
OS
XX

```

```

PN      W0200157190-A2.
XX
XX      09-AUG-2001.
PD
XX
XX      05-FEB-2001; 2001WO-US04098.
PE
XX
PR      03-FEB-2000; 2000US-0469914.
PR      27-APR-2000; 2000US-0560875.
PR      20-JUN-2000; 2000US-0598075.
PR      19-JUL-2000; 2000US-0620325.
PR      01-SEP-2000; 2000US-0654936.
PR      15-SEP-2000; 2000US-0663561.
PR      20-OCT-2000; 2000US-0693325.
PR      30-NOV-2000; 2000US-0728422.
XX
XX      (HYSE-) HYSEQ INC.
XX
XX      Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI      Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI      Xue AJ, Yang Y, Wejhrman T, Goodrich R;
DR      WPI; 2001-476283/51.
DR      N-PSDB; AAK51958.
XX
XX      Nucleic acids encoding polypeptides with cytokine-like activities,
PT      useful in diagnosis and gene therapy -
XX
XX      Claim 20; Page 3768-3770; 6221pp; English.
XX
XX      The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC      encoded polypeptides (AAAM78323-AAAM80302) that exhibit activity elating to
CC      cytokine, cell proliferation or cell differentiation or which may induce
CC      production of other cytokines in other cell populations. The
CC      polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC      peptide therapy. The polypeptides have various cytokine-like activities,
CC      e.g. stem cell growth factor activity, haematopoiesis regulating
CC      activity, tissue growth factor activity, immunomodulatory activity and
CC      activin/inhibin activity and may be useful in the diagnosis and/or
CC      treatment of cancer, leukemia, nervous system disorders, arthritis and
CC      inflammation.
CC      Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC      (AAAM80020) are omitted as the relevant pages from the sequence listing
CC      were missing at the time of publication.
XX
XX      Sequence 1026 AA:
SQ

```

Query Match 4.5%; Score 335.5; DB 22; Length 1026;
Best Local Similarity 21.8%; Pred. No. 1e-11;
Matches 189; Conservative 142; Mismatches 341; Indels 193; Gaps 38;

```

QY      543 RSRSPRYRYSRGRS-PPEFRGOSPTKRNVPREKEKEYNRNKREVPDPIDIRAYRS 601
Db      312 rsageelteryqlqartlealeakstkdslerqsele--drbqa-----di-asyqea 363
QY      602 VDFRPPFEKERYREWE-----RKYREW-----YEKKYG--YAVGQPRSA 641
Db      364 igqld--aelnrtkwemaaglireyqdlinvkmaaldiaayfklllegeccrtigfipr-- 419
QY      642 NREDFSPERLLPLNIRNSPTFRGRRED--YAAGOSHRRNRNLCGNYPEKLISTRDSHNADN 699
Db      420 ----tslpegiplkipsvthlkvkseeklkvkekekativveeqtelqyleveetee 475
QY      700 PKSKKESEBNVPGDGKGNKHKHKRRRNEKEGESESLNPPELLFTSRCKGSSSIDETK 759
Db      476 keakeeg-----keeggeee-----eggeetk 502
QY      760 TDTL--FLVLPSSDDATPVRE---PMDAESTFKVSVDKDKREKDKPVKVSUKTRKSDG 814
Db      503 spvveaaspeakeaspkveekspakeas-----pekeekaspeevkspeaksp 554
QY      815 SATAK-----KDNVLKPSKGPOEKVDGDRKSPRSEPLKRAKEBATKIDSVKPSSSQK 869

```


QY 1222 LSEGGHFKTLSSQSKETRTSEKHESVYRGSSNKDFTPGRDKKVYD-----DSRYSSSSKRRD 1277
Db 1058 alekge---aeqseeeadeedkaedar-----eeeyepekeaaedyamavvdkaaeagaae 1111
QY 1278 ENGELARRRDSPPR--GKSLSLGQSKSLKEERDLPPKGAESKSSNSPP-RDKKPHDKA 1334
Db 1112 qygfilt---tpkqlgagspgprepasslhdelp-ggsesatlasdeenredqpeefta 1166
QY 1335 ----PEETKPCDEETKPVVDKNSGKEREKHAAREARNGKSSGANCXYLMR 1380
Db 1167 tsyqstleisseptmdems-tpdrmsdetmneetespeqefvnlck 1215

RESULT 15
ABBB1598
ID ABBB1598 standard: Protein; 1151 AA.
XX ABBB1598;
AC ABBB1598;
XX 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 11586.
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical.
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL05701.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PS
XX Disclosure; SEQ ID NO 11586; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB557737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1151 AA;

Query Match 4.5%; Score 332.5; DB 22; Length 1151;
Best Local Similarity 19.6%; Pred. No. 1.8e-11;
Matches 268; Conservative 173; Mismatches 459; Indels 465; Gaps 57;

QY 84 ECRRTLLSSDKHTCTPCGQNDVSPDALIANFLROAVN---NFKMETGYTKLRKQLPP 140
Db 67 erlrkhmdsd-----vldkygdvdqdvnpkvatlfepkieslyvk 109

QY 141 FLFLVPPPP--LSQRNLQP-----RSRSPI-----LRQODPVVERVTVSPIC 181
Db 110 yigtipparptmlsppllppfigahngssllnveltvglllpdlqispsdratrvkss 169
QY 182 SPTKTAGSCSDSGYLSRLPAPSTISLTSNQSSLAPVSGNPSAPAPVDITATVTSVH 241
Db 170 rdelldpyddedt-----spsy-elvsertkltaleeinnvs-----lmsdsvngv 216
QY 242 SEKSDCPFFDSQNKLLPAAALTS-----EHSKGASITATLMEKGV--GTSPPNSLIF 294
Db 217 sqasqlsgvssdrllmaaslesmvdshqhtaahnsgepenlceeqmpekfsensvds--- 273
QY 295 VGSSTLHGQLIPTGPVRINAAAPGSGR-----PGWHSNKLGLYVSP 337
Db 274 -----asagfgrsegqrlhdfkdqdaiftfgterkhsavsetlt 312
QY 338 PQQIRRGF---RSCYRSINCRGHHNSQRTQSPSLPATPCFVPPV---PLYPPP 389
Db 313 glqlslsiedelnsevkkanidannasieslapdevgvpapkvnppppprpsapkvpp 372
QY 390 HTLPPLPGV-----PPQPSQPF-----PSSOPTAGYSVPPGFPAPANIST 433
Db 373 ---pappgyesppgpqppasprfdppphltieppppppl---vppp--ppapevepp--- 422
QY 434 ACFSPGVPTAHSNTMPTTQAPLISREEFYEQNDKGRESKFPYSGSSYSTDSQGL 493
Db 423 ---kpppppa-----pptvpp----- 436
QY 494 AQHIALTLSPSAHTLDLHDHPPEEAARSAMIVHMPDLMIDIAHRSPPRYR 553
Db 437 -----pppppaprvt-----epppppppaprkt---veppp--ppapevepp--- 474
QY 554 SRSRSPPEFGOSPTKRNVPREKEREYFNRYREVPDPDIAFYRSVDFRDPFEKRY 613
Db 475 ---pp-----aptelepp-----pppaprkt---velppp----- 497
QY 614 REWERKRYREWEKYYKGAVGAQPRPSANREDSPRELLPLNIRNSPTFRGRREDYAAG 673
Db 498 -----pappkaeaaltpr-----pappkaeaaltpr-----akgs 515
QY 674 SHRNRNLGNYPPEKLISTRDS-HNAKDNPKSKESESNVPGDGKGNKHKRRKRNEKE 732
Db 516 ngfteleavt-pkesetrdkphvnlvdvkedepa-----ttedkqd 556
QY 733 ESESFANPELLERSKRCRSSGIDETKTDTLVLPBRDQATPYRDEPMAESTTFESVD 792
Db 557 kp-----lel-----gladap-----davstasestpslt-ssgsk 588
QY 793 KDRREKDKPKVKSDKTRRKSDGSATAKDNVLKPSKPOEKVGD---REKSPRESEPL 848
Db 589 shskstdkekekdrrhnrhd-----dkhrrtsldrdtdrdrskshss 635
QY 849 KKAKEBATKIDVYKPSSSQKDEKVTGTPRKAHKSAGKDTROSQPRTRRSKRTVPTSS 908
Db 636 ssakhshssssskhshsskndksssssssrnsressskrgtsssrheesshkhkh 695
QY 909 QKSQPVTRPRSLKINVLIAEKNEKREKRSYVUKDESSSMKISKVGEITVAPSK 968
Db 696 sssssssseidkykek-----dkereksqgrshssssssss-----sar 735
QY 969 RKMEGVEKLETPPEKDKIASSTTPAKTIKL-----NRETOKKIGNAENASTKEPS- 1020
Db 736 rkdh-----dgrdrdrnkstsgsaenkalhdhsekekykqrrgsdndegkppss 789
QY 1021 -----EKLSTSSKIKQEKVKGKARKVAGSESSSTLVD-----YTSTSS- 1061
Db 790 gypaknsqpedsaatknsdapenangtngnsngstngcadvngvylsdlqgtsf 849
QY 1062 ---TGSSPVKRSSEKEDTKRTVYKTMEEYNNNDTAAEDVYIMIVQSKWKDKDPFESSE 1118
Db 850 veltags---gsidreasskhepedldgkeadqp-----ektelearg 890
QY 1119 EDVKTTQPIQSVGKPPSSIIKNTTKPSATAKYT-----EKESQPEKIQGLPKE 1167

Search completed: September 12, 2002, 18:28:42
Job time: 11242 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 16:03:56 : Search time 25.89 seconds
(without alignments)
1324.586 Million cell updates/sec

Title: US-09-811-045A-1
Perfect score: 7374
Sequence: 1 MWEKDPNMGKMLNTGKY.....WRSWLGRWRPSSRNP 1404

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCPUS.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1 | 340 | 4.6 | 3969 | 4 US-08-061-376-5 | Sequence 5, Appl |
| 2 | 276.5 | 3.7 | 1187 | 1 US-08-320-559-28 | Sequence 28, Appl |
| 3 | 276.5 | 3.7 | 1187 | 3 US-08-545-860D-28 | Sequence 28, Appl |
| 4 | 276.5 | 3.7 | 1187 | 5 PCT-US94-04496-28 | Sequence 28, Appl |
| 5 | 276.5 | 3.7 | 1210 | 1 US-08-320-559-26 | Sequence 26, Appl |
| 6 | 276.5 | 3.7 | 1210 | 3 US-08-345-860D-26 | Sequence 26, Appl |
| 7 | 276.5 | 3.7 | 1210 | 5 PCT-US94-04496-26 | Sequence 26, Appl |
| 8 | 274 | 3.7 | 2842 | 1 US-07-741-940-7 | Sequence 7, Appl |
| 9 | 274 | 3.7 | 2842 | 1 US-08-289-548A-7 | Sequence 7, Appl |
| 10 | 274 | 3.7 | 2842 | 1 US-08-452-654-7 | Sequence 7, Appl |
| 11 | 274 | 3.7 | 2843 | 1 US-08-452-655B-2 | Sequence 2, Appl |
| 12 | 274 | 3.7 | 2843 | 1 US-08-452-655B-7 | Sequence 2, Appl |
| 13 | 274 | 3.7 | 2843 | 3 US-08-450-582-2 | Sequence 2, Appl |
| 14 | 274 | 3.7 | 2843 | 3 US-08-450-582-7 | Sequence 2, Appl |
| 15 | 274 | 3.7 | 2973 | 2 US-08-821-355A-7 | Sequence 7, Appl |
| 16 | 274 | 3.7 | 2973 | 2 US-09-003-687A-7 | Sequence 7, Appl |
| 17 | 274 | 3.7 | 2973 | 4 US-09-136-605-7 | Sequence 7, Appl |
| 18 | 273.5 | 3.7 | 1588 | 5 PCT-US93-07261-11 | Sequence 11, Appl |
| 19 | 273.5 | 3.7 | 1663 | 5 PCT-US93-07261-16 | Sequence 16, Appl |
| 20 | 273.5 | 3.7 | 1780 | 1 US-08-769-309A-5 | Sequence 5, Appl |
| 21 | 273.5 | 3.7 | 1780 | 1 US-08-994-570-5 | Sequence 5, Appl |
| 22 | 272.5 | 3.7 | 1162 | 2 US-08-728-323A-2 | Sequence 2, Appl |
| 23 | 270 | 3.7 | 2843 | 1 US-07-741-940-2 | Sequence 2, Appl |
| 24 | 270 | 3.7 | 2843 | 1 US-08-289-548A-2 | Sequence 2, Appl |
| 25 | 270 | 3.7 | 2843 | 1 US-08-452-654-2 | Sequence 2, Appl |
| 26 | 270 | 3.7 | 2843 | 1 US-08-370-235A-2 | Sequence 2, Appl |
| 27 | 264 | 3.6 | 688 | 3 US-09-141-047-8 | Sequence 8, Appl |

| | | | | | |
|----|-------|-----|------|--------------------|-------------------|
| 28 | 262 | 3.6 | 1898 | 1 US-08-056-200-94 | Sequence 94, Appl |
| 29 | 262 | 3.6 | 1898 | 2 US-08-800-644-94 | Sequence 94, Appl |
| 30 | 261.5 | 3.5 | 1581 | 4 US-09-110-517-2 | Sequence 2, Appl |
| 31 | 259 | 3.5 | 434 | 1 US-08-097-830E-3 | Sequence 3, Appl |
| 32 | 259 | 3.5 | 434 | 2 US-08-456-112B-3 | Sequence 3, Appl |
| 33 | 250.5 | 3.4 | 1706 | 2 US-08-459-568-2 | Sequence 2, Appl |
| 34 | 250.5 | 3.4 | 1706 | 2 US-08-399-411-2 | Sequence 2, Appl |
| 35 | 250.5 | 3.4 | 1706 | 2 US-08-516-859A-2 | Sequence 2, Appl |
| 36 | 241 | 3.3 | 1400 | 1 US-08-080-255-7 | Sequence 7, Appl |
| 37 | 241 | 3.3 | 1400 | 1 US-08-465-713-7 | Sequence 7, Appl |
| 38 | 241 | 3.3 | 1400 | 5 PCT-US93-05857-7 | Sequence 7, Appl |
| 39 | 235 | 3.2 | 683 | 6 5210183-3 | Sequence 7, Appl |
| 40 | 226 | 3.1 | 885 | 2 US-08-533-306A-4 | Sequence 4, Appl |
| 41 | 226 | 3.1 | 885 | 2 US-08-742-923A-4 | Sequence 4, Appl |
| 42 | 225 | 3.1 | 1719 | 2 US-08-459-568-4 | Sequence 4, Appl |
| 43 | 225 | 3.1 | 1719 | 2 US-08-399-411-4 | Sequence 4, Appl |
| 44 | 225 | 3.1 | 1719 | 3 US-08-516-859A-4 | Sequence 4, Appl |
| 45 | 224 | 3.0 | 571 | 4 US-08-961-083-4 | Sequence 4, Appl |

ALIGNMENTS

RESULT 1
US-08-061-376-5
; Sequence 5, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Djabbal, Malek
; APPLICANT: Seller, Licia
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11023
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; City: Los Angeles
; STATE: California
; COUNTRY: USA
; Zip: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)546-4737
; TELEFAX: (619)546-9392
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3969 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-061-376-5

Query Match 4.6%: Score 340; DB 4; Length 3969;
Best Local Similarity 20.1%: Pred. No. 9.5e-14;
Matches 283; Conservative 153; Mismatches 463; Indels 512; Gaps 60;
QY 111 LIANKFLQAAVNNKFNKGTGTRLRKQLPP-----LFLVPPRPPLSRNLQ 157

```

Db 370 LLSNRYGQ-----KKGAQKKEKAAOLOGRKVTQYKNIROFIMPVSAISSRIK 422
QY 158 PRSRPIRLQODPV-----FRYVSPICSDKTKTGSCSD-----SGLSLRLAP 202
Db 423 TPRRFIEDEYDPPKIALESTPNRHS-APSCGSEKSSAASQHSQMSQSSRSSSP 441
QY 203 SISLTSNOS-----LAPVSGNPSPAPVPDITATVSIYHSEKSDPFRDSNKL 257
Db 482 SVDTSDQASAEIOVLPRESDTPEVHP-PLP-----ISQSPENESNDRSRHS- 531
QY 258 PAALISEHKGASIAITAIAMEKGVPGTSPMNSITFGOSLHGOILPTTGVRINAR 317
Db 532 -----VSESRFGSRT----- 541
QY 318 PGGGRPGMEHNSKLYTVSPPOOIRGERSCTYRINNGRHRSOROTOSPLPATPCFV 377
Db 542 -----TKLSTLOAPQO-----QTSSP----- 560
QY 378 PVPPPLPYPP-----HTLPLPGVP--POFSPQFPSSQPTAGS---VPPPG 423
Db 561 --PPPLTTPPPLOPAPASSIDHT--PWLMPPTILASPPLPASTAPMOKRKSILREPT 615
QY 424 F-----PPRANISTACFS-PCV--PRAHNTMPTTAPILSREPTREONDGRES 472
Db 616 FRMTSLKHSRSEPOYFSSAKYAKEGLRKPFDNFRPPLTP-----EDVGFAS 664
QY 473 KFPYSGSSYRSTYTSSOGIAOHIALTLSPAHTLIDLHDPHPPEAEARSAMIVH 532
Db 665 GFASAGTAAARLPLSHSTGRDMH--KSP-----LLRAPRTPSAHSRIEVSPT 715
QY 533 MDPLMDIAHARSPPYRRTSRSSRSPFERGOSPTKRNVPREKEREYFNRYREVPY 592
Db 716 LPSNRTSAGTSSGVSNRKRKRYFSPIRSEPRSP--HSMRTSGLSSELPLTPPS 773
QY 593 DIRAYYGRVD-----FDPEEKERYREMERKYREMYEKGAVGQAPR-- 639
Db 774 SVSSSISIVSPATSAALNPTFFPS-----HSLTOSGEAEKQNRKQ 818
QY 640 -SANREDF--SPERLLPLNIRNSPPTGRREDYAGQSHRNRLNGNYPEKLS-PROSH 694
Db 819 TSAAPRPFSSSPPLPFWPTPSSOTERGKNAKA-----PEELSKDRAD 864
QY 695 NAKDNPKSKRESENVPGDGKGNKHKRRRMEKEGESEFLNP-----ELL 743
Db 865 KVEKDKSREDRER--EKENKRESRKEKRGKSIEQSSALYPGVRYSKEKVEDV 920
QY 744 ETSRKGCGSSGIDETKDTLFLVLRSDDATPVADDEPMDESTIFKVSQDKREKDKPY 803
Db 921 ATSSSAKKAIGRKKSS-----SHDSGT-----DITSVTL----- 949
QY 804 KSDTKRKSDGSAKAKDNVL--KPSKPOEKVDGREKSPRSEPLKAKE--EATKID 859
Db 950 -----GDTTAVTKTILIKKGRGNLEKTLID--LGP-TAPSLKEKTKLCLSPSS 996
QY 860 SVKPSSSS-----QKDEKVTGTPRKASHSKAKDTRROSPRTFRSKRYTPKTS 907
Db 997 TVKHTSSISGMLAQADKLPMTDKRVASLKKAKAQOLCKEKSKSLKQTDQ-----PKAQ 1051
QY 908 SQKSQVPR--RRPSRLKIKNTILIAEKNKEREKRSVDKDFESSMSKISKVEGTIVK 965
Db 1052 GQSDSSESTSVRGR--IKH-VCRRAVALGRKRAVFPD----- 1087
QY 966 SPRKMEGDEVKELTRP--EKDKIASSTTPAKKIKLNRGKKIGAENA NSTKEPSEKL 1023
Db 1088 -----DMPITSLALPWEERKILSS-----MGNDKSSIAGEDEPIAPIKPI 1131
QY 1024 ES-TSSKIKOEK--VKGAKRKVAGSEG-----SSSTLVDTYSTSGSPVSKSEKTD 1075
Db 1132 KPYTRKKAQOEPPVKKGRSGRGCGCPGCVCTNCIDKPKFGGNIKKQOCK-- 1189
QY 1076 TKRTVKTMEEYINNDTAPAEVYIIMIQVPOSKWMDKDFESEEDVKTQPIQSVKPS 1135

```

```

Db 1190 -----NRKCONLQMPESKAYLOQANAVKKREKSKTSEKDD-----SKESS 1231
QY 1136 IIRKNTTKSATAKYTEKESEOEPEKLOKLPKEASHLQHEILSSKGSASSEKGRAKRE 1195
Db 1232 VKNVNV-----DSSQKP----- 1243
QY 1196 HSGSEKDNPKRKSQAOPKKESTVDRLSEQHFRTLQSSKETRTSEKHEVSGSSNKDF 1255
Db 1244 -TPSAREDPAPKKSSEPPPKRVEEKESEGNVSA GPESKQATTPASRKSQVQOPAL 1302
QY 1256 -----TGORDKVVYDSDYSSSKRDERGELARKDSPRGKESLSGQSKLREED 1308
Db 1303 VDPDPPPTGPPRK-----EVPTTSEPKKQD-----P 1332
QY 1309 LPKGAESKKSNSPPRDKPHDKAPYETKRPCEETKPDK-----NS 1352
Db 1333 PPSGEPQSKQKAVAPRSPVPKQ--PKKEKP-----PPVKNQENAGTILSTLSNGNS 1387
QY 1353 GKER-----EKHAARNGKESSG 1371
Db 1388 SKQKIPADGVHRIYDFKEDCEAEVWEMMG 1418

```

RESULT 2

```

US-08-320-559-28
; Sequence 28, Application US/08320559
; Patent No. 5633135
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135rls
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320, 559
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,094
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-0855
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 amino acids

```

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-320-559-28

Query Match 3.7%; Score 276.5; DB 1; Length 1187;
Best Local Similarity 18.3%; Pred. No. 3.2e-10;
Matches 227; Conservative 160; Mismatches 428; Indels 423; Gaps 55;

QY 231 DITATVSVISVHSEKSDGFRSDNKL-LPAALITSEHSKGASSIAITALEMEKGVPGTSP 289
DB 43 EVKEFLSTKSHTRLDA-----SENRLGPKYPLIPD-KGSS-----IPSSSF 84
QY 290 WNSIVVGSLSLHGOLI-PTTGPVRIN--AARPGGGRPGWE-----HSNKLCYLVSPQ- 339
DB 85 HTSV-----HHQSHHTPASGPLSVGNISHPKMAQPRTEPMPSLHAKSCG-----PPDS 133
QY 340 ----QIRGERSCYRINRGHHSERQRTQSPSLPATPCFVPVPPPLYPPTHLP 395
DB 134 CHLTODRLGQEGFGSS-----HHKGDRAAG-----DHCAST 167
QY 396 PGVPPPOESP--QFPSSQPTAGYVPPGPPAPANISTACSPGVPTAHSN--TMPTT 451
DB 168 DSAPFERELSLTSLSPVPP-----LSP-----HSHNQTLPT 201
QY 452 QAPLISREFIREQNDKRGESKFPYSSGSSYSRST--TDSSGGLAOHII-----A 499
DB 202 QG-----SSK--VHGSSNNKGVCYCPAKSPKDLAVKVDKETPODSLVA 242
QY 500 LTLSPSAHTLDLHDHPHPEAEARSAMIVHMDLMDIAHARSPPRYRYSRSRSP 559
DB 243 PAQPS-----QTFPPSLPSKSYAMQKP-----TAYVRMD-----GQDAAP 281
QY 560 PEFQSGPTKRNVPREKEREYFNRYREV-----PPYDIKAYYGRSVDRDPE 609
DB 282 SE-----SEELKPLPDYQOQTEKTDLYKPAKAKLTLLKMPQSQVEQYTSNEVHCVEITL 337
QY 610 KERYEMERKRYEMYEKYYKGYAQAQPPPSANREDFSERLPLNIR-----SPTRR 665
DB 338 KEMTHSW-----PPLTALHTPTSTAPSKFPPTPKDSQHVSVTQMO 379
QY 666 REDYAAGOSHRNRNLGN--YPEKLISTRDSHNAKNPKSEKESFNVPDGGKNNKHKHR 723
DB 380 KQYDTSKTHSHSQGTSMLLEDJOLSD-----SEDSDSQTP----- 418
QY 724 KRANEKEBESFLNPELL-ETSRCKGSSGIDETKTDTLFLVLRSDATPVDRPMDA 782
DB 419 ----EKPPSSSAPPSAPQSLPEPVASAHSSSAESESSTDS-----DSSSDS 460
QY 783 ESTFKSVSDKK-REKDKPKYKSDKTRKKSQDSATAKKDNVLRKSGQEKVDDREKS 841
DB 461 ESESSSSSEENPLETTPAPEPEPTTNKWOJDMNLTKSQPAAPPEGRS-----TEP 514
QY 842 PRSEPLPKAKEEATKIDSVKPPSSSQKDEKVTGTPRAKHSANQDTRROSQPTRRSKR 901
DB 515 PRHRESKSSSAT-----SQEHSKSDPPKSSSKAPAPAPAPHP-----GKR 560
QY 902 TVPKTSSQKQVPR-----TRPRSLKINYLIAREKNEREKKKSVKDKDESSSKISKIV 957
DB 561 SQKSPADQEPQRTQVTKQPKPKPVK-----ASAAAGSRSTSLOGREREGLLPYGSR 612
QY 958 ECTELVKKSPKMKMEGDVEKLEKTEPEKDKIASSTTPAKKIKINRELGKIGNAENASTJK 1017
DB 613 DOTS--KDKPKYKTG-----RPRAAASN 634
QY 1018 EPSEKLESTSTKIQOEKYGKAKRRVAGSEGSSTLVDT-----STSGSGSPVRKS 1070
DB 635 EKPAVPSPSEKSKKKHSSSLPAPSKALSGEPAKDNVEDTPEHFALVPLTEESOGPRHSS 694
QY 1071 EKETPTKRVITMEBYNNNDNTAPADVYIMIQVPOSKWKDKDFSESEEDVYTTQPTQSV 1130
DB 695 GSRTSGCRQAVVVOEDSRKDR-----LPLPL-----RDTKLISPLRDT 732

QY 1131 GKPSSTINVTTPKPSATAKYTEKESEQPEKLOKLPKEASHELMOHELASSKGSASSEKGR 1190
DB 733 PPOSLMWKITL-----DLISRIOP-----PGKG---SRQK 762
QY 1191 AKDRHSGSEKDNPDKRRSGAOPDKESTYVDRLSEQGHFKTLSQSKETRTESEHBSVGS 1250
DB 763 AEDKPPAGKHHSEKRS-----SDSSKLAKRKGAEKEDC 799
QY 1251 SNKDETPGRDKKVDYDSRDYSSSKRRDERGELAR-----RKDSPRKEKESLSQK---- 1300
DB 800 DKKIR--LEKETKQSSSSSSSHKESKTKPSRPSQSSKEMLPPEPVSSSQKPAKP 857
QY 1301 --SKIREEDL-----PKGAEKSN--SSPPDK-----PHDK-APYETKRP- 1341
DB 858 ALKRSRRADTCGODPPKRSASSTKSNHKDSIPIQGRVREGKGRSSSEHKSSGDTANPF 917
QY 1342 -----CEETKPVDNKSGKEREKHAHAENGRESS 1370
DB 918 PVPSPNGNSKRGKPYQKFDKQADLHREAKKMKQKA 955

RESULT 3
US-08-545-860D-28
Sequence 28, Application US/08545860D
Patent No. 6040140
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: CanaanI, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1 Region
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
ADDRESSEE: No. 6040140ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545, 860D
FILING DATE: 07-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496
FILING DATE: 22-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10930
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/327,392
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/320,559
FILING DATE: 11-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,443
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,094
FILING DATE: 30-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,839
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/805,093

```

1 BILLING DATE: 11-DEC-1991
2 ATTORNEY/AGENT INFORMATION:
3 NAME: Deluca Esq., Mark
4 REGISTRATION NUMBER: 33,229
5 REFERENCE/DOCKET NUMBER: TJU-1262
6 TELECOMMUNICATION INFORMATION:
7 TELEPHONE: (215) 568-3100
8 TELEFAX: (215) 568-3439
9 INFORMATION FOR SEQ ID NO: 28:
10 SEQUENCE CHARACTERISTICS:
11 LENGTH: 1187 amino acids
12 TYPE: amino acid
13 TOPOLOGY: linear
14 MOLECULE TYPE: protein
15 OS-08-545-860D-28

```

| | | | | | | | |
|-----------------------|--------|--------------|--------------|------------|------|--------|-------|
| Query Match | 3.7%; | Score | 276.5; | DB | 3; | Length | 1187; |
| Best Local Similarity | 18.3%; | Pred. | No. 3.2e-10; | | | | |
| Matches | 227; | Conservative | 160; | Mismatches | 428; | Indels | 423; |
| | | | | | | Gaps | 55 |

| | | | |
|----|-----|--|-----|
| Oy | 231 | DITTVASJVSJHSEKSDSDFRSDUKL-LPAAALJHSEKSSASSIATVJALMEGVJGJVSF | 269 |
| Db | 43 | EVEKEELSTKSHTHLDA-----SENRLKPKYPLIPD--KGSS-----IPSSSF | 84 |
| Oy | 290 | WNSJFVGOSLHGOLI--PTTGPRVIR--AARPGGGRPGME-----HSNKLGLVSPQ- | 339 |
| Db | 85 | HTSV-----HHOSIHTPASGPLSVGNISHPNMAOPRTPEMPSLHASCQ-----PPDS | 133 |
| Oy | 340 | -----OIRGERSGCRSTNRGNHHSERSGORTQSPSLPATPCFVFPVPPPLYPPTPHLPLP | 395 |
| Db | 134 | OHLYODRIGQGFQSS-----HHKKGRADG-----DHCASTY | 167 |
| Oy | 396 | PGVPPQPSF--QEPSSQPPAGSVPPPGFPAPANISTACEPSGVPAHSN--TMPTT | 451 |
| Db | 168 | DSAPRELSPLLSLSPVPP-----LSP-----IHSNOQLTPRT | 201 |
| Oy | 452 | QAPLLSREEFYREONDGRESKFPYSGSSYSRSSY--TDSSQGLAOIH-----A | 499 |
| Db | 202 | QC-----SSK--VHGSSNNKCYCAKSPKDLAVYVHNKETPODSLVA | 242 |
| Oy | 500 | LTLSSAAHTLDLHDHPPEBEAERASAMLVNHPDLMJAHARSRSPPRYRSTRSRSP | 559 |
| Db | 243 | PAOPPS-----QTFPPPSLSPSKVAQAOKP--TAYRPMD-----GDDQAP | 281 |
| Oy | 560 | PEFRQSPTRKRVPREEKEREYFNRYREVP-----PYDIKAYYGRSYDFRDPFE | 609 |
| Db | 282 | SE-----SPLAKLPEDDYQOQFTEKTDLVPAKAKLTUKMKMPSQSEVQTSNEVHCVEIL | 337 |
| Oy | 610 | KERYREWEKRYREWEKYYKGYAVGAQPRPSANREDSPEFLPLNIRN-----SPTFRGR | 665 |
| Db | 338 | KEMTWSM-----PPLTLTAIHTPSTAEPSKFPFPKDDSOHVSVYQO | 379 |
| Oy | 666 | REDYAGAGOSHRRKRNIGCN--YPEKJSTDSINADNPASKKESEENVPGDGKGNKKHHR | 723 |
| Db | 380 | KOYDSSSKTHNSOOGTSMLEDDJQISD-----SEDSDEQTP-----418 | |
| Oy | 724 | KRNEKEGESESPLINPELL--ETSRKCGSSGIDETKTDTLFVLPSRHDATPYADEPMA | 782 |
| Db | 419 | ----EKRPSSSAPPAPOSPLRPVAVASIHSSASSSESTSDS-----DSSSDS | 466 |
| Oy | 783 | ESITTFKSVSDMK-REKDKPKVXSQTKRKSDGSATKAKONVLPKSGPOGEKVDGPREKS | 841 |
| Db | 461 | ESSSSSSDEENEPLETPAPEPEPTTNKWOJNLWLTQVQAPAPRQSPRS-----TEP | 514 |
| Oy | 842 | PNSEPLLAKAEATKIDSVKPSSSSQKDEKVTGTPRKAHKSASAKDTRROSQPTTRSKR | 901 |
| Db | 515 | PRNHPESAKSSDSAT-----SOHSSEKQPPRPSSSSKAPRAPPEAPHF--GKR | 560 |
| Oy | 902 | TVPKTSSQKQVVR-----TRPRSLRKJNVLAREKNREKRRKSVYDQFSSSMKISKY | 957 |
| Db | 561 | SOCKSPAOGEPOROTGVTKQOKKPVK-----ASARAGSRTISLOGEREPGLLPGR | 612 |

| | | | |
|----|------|---|-------------------------|
| QY | 958 | EGEIEIYKPSKRMEDVEDKLEERTPEKOKIASSTPAKKIKLINFETGKIGMAENASTTK | 1017 |
| Db | 613 | DOYS--KOKKXATTKG----- | PRRAASN 634 |
| QY | 1018 | EPESEKLESTSSKIKOEKVKGAKRKAAYAGSEGSSTLYDYT----- | STSTGSGPVAKS 1070 |
| Db | 635 | EPRKAVPSPSEKKKKHNSLPAPSKALSGEPRAKDVNEDRTPEHALVPLDTEOGGPHSGS | 694 |
| QY | 1071 | EKTDTKRYTIKIMEEYNDNNTAPADYILIMIQVOPSKMDKODFSEEDVYKTPPISV | 1130 |
| Db | 695 | GSRTSGCROAVVVOEBSRKDR-----LPPLP----- | ROTKLPSLMDT 732 |
| QY | 1131 | GKPSIIKNTYTPSPATAKYTEKESEOPKLOKLPKASHELMOHELNSKGSASSEGR | 1190 |
| Db | 733 | PPPSLWVKITL-----DLSPRIQP----- | PGKE---SRORK 762 |
| QY | 1191 | AKDRHSGSEKDNPKDRAKSGAQDPESTYDRLSEOGHFTLSQSSKETPTSKHESVGS | 1250 |
| Db | 763 | AEDDQAPAGKAKHSEKSS----- | SDSSSKLAKKKGAEBCD 799 |
| QY | 1251 | SNKQFTGPRKKVYDYDRDYSSSKRRDERGELAR----- | RKDSPPRGESLSGOK--- 1300 |
| Db | 800 | DNKXIR--LEKEIKSGSSSSSSSHKESKTKPRPSSQSKKEMLPPPVSSSSQKPAKP | 857 |
| QY | 1301 | --SKLBERDL---PKGAESKKS---SSPRDK----- | PHDK-APYETKRP- 134 |
| Db | 858 | ALKRSRRADTCQGDPPKASSTKSNHKDSSIPKORRVLEGKGRSSSEHKGSSGDTANPE | 917 |
| QY | 1342 | -----CEFTKPVDNKSGSKEEKKHAARNGKSSS | 1370 |
| Db | 918 | PVPSPLPNGNSKPGRPQYKQQAQADLHREAKKMKOKA | 955 |

```

1      RESULT      4
2      PCT-US94-04496-28
3
4      ; Sequence 28. Application PC/TUS9404496
5
6      ; GENERAL INFORMATION:
7
8      ; APPLICANT: Croce, Carlo
9
10     ; APPLICANT: Canaan, Eli
11
12     ; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
13
14     ; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
15
16     ; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1-
17
18     ; NUMBER OF SEQUENCES: 86
19
20     ; CORRESPONDENCE ADDRESS:
21
22     ; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
23
24     ; ADDRESSEE: Norris
25
26     ; STREET: One Liberty Place, 46th floor
27
28     ; CITY: Philadelphia
29
30     ; STATE: Pennsylvania
31
32     ; COUNTRY: USA
33
34     ; ZIP: 19103
35
36     ; COMPUTER READABLE FORM:
37
38     ; MEDIUM TYPE: Floppy disk
39
40     ; COMPUTER: IBM PC compatible
41
42     ; OPERATING SYSTEM: PC-DOS/MS-DOS
43
44     ; SOFTWARE: Patent In Release #1.0, Version #1.25
45
46     ; CURRENT APPLICATION DATA:
47
48     ; APPLICATION NUMBER: PCT/US94/04496
49
50     ; FILING DATE:
51
52     ; CLASSIFICATION:
53
54     ; ATTORNEY/AGENT INFORMATION:
55
56     ; NAME: Deluca Esq., Mark
57
58     ; REGISTRATION NUMBER: 33,229
59
60     ; REFERENCE/DOCKET NUMBER: TJU-1242
61
62     ; TELECOMMUNICATION INFORMATION:
63
64     ; TELEPHONE: (215) 568-3100
65
66     ; TELEFAX: (215) 568-4439
67
68     ; INFORMATION FOR SEQ ID NO: 28:
69
70     ; SEQUENCE CHARACTERISTICS:
71
72     ; LENGTH: 1187 amino acids
73
74     ; TYPE: amino acid
75
76     ; TOPOLOGY: linear
77
78     ; MOLECULE TYPE: protein
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```


PCT-US94-04496-28

| | | | | |
|-----------------------|-------------------|--------------------|-------------|--------------|
| Query Match | 3.7%; | Score 276.5; | DB 5; | Length 1187; |
| Best Local Similarity | 18.3%; | Pred. No. 3.2e-10; | | |
| Matches 227; | Conservative 160; | Mismatches 428; | Indels 423; | Gaps 55 |

| | | | |
|----|------|--|------|
| QY | 231 | DITVTSIVSHVSEKSGOPFRDSDNKT-LPAALITSHSGASGASTITAMLEEKVYGPSP | 289 |
| Db | 43 | EVEFELSTKSHTRDLA-----SENRLGKKRYLIPD--KGSS-----IPSSSF | 84 |
| QY | 290 | WNSLIEVGOSLLHGOLI-PTTGVRIN--AARGGGRPWE-----HSNKLGVLVSPQ- | 339 |
| Db | 85 | HTSV-----HHQSHITPASCPLSGVGNISHNPKMAQPTPEMPSLHAKSG-----PPDS | 133 |
| QY | 340 | -----QIRGERSGVCYRINRHRHSHSESOQTQSPSLPATPCFVVPVPPPLPPPTLTLP | 395 |
| Db | 134 | OHLTQDMLGEGGSS-----HHKKDRKRDG-----DHCASTV | 167 |
| QY | 396 | PGVPPQSP--QFSPSQPTAGVSVPPGFPAPANISTACSPGVTAHSN--TMPTT | 451 |
| Db | 168 | DSAPERLSPILSPVPP-----LSP-----HSHQOTLPT | 201 |
| QY | 452 | QAPLISHEEYRREONDKGRSKFPYSGSSYSKSY--IDSSGLAQHIN-----A | 499 |
| Db | 202 | QG-----SSK--VHGSSNNSKGCYCPAKSPKDLVVKHDKETPODLSVA | 242 |
| QY | 500 | LTLSPSAHTLDLHDHPHPEEAEARSAMIYMPDLMIANARSRSPRYRRSRSP | 559 |
| Db | 243 | PAQPPS-----QTFRPPLSPKSVYAMQKP--TAVVRMD-----QGDAP | 281 |
| QY | 560 | PERFGOSPTKRWPREKEFEYNNKRYEV-----PPYDIKAYGSRVDFRDFE | 609 |
| Db | 282 | SE-----SPELKPREDYRQOTFEKTDLKYAPAKLTKLMPQSOVEQYUSMEHCEVETL | 337 |
| QY | 610 | KERYREMYRYREMYEKYGVAVGAQPPSARREDPSEPLPLNIRN---SPETGR | 665 |
| Db | 338 | KEMHSM-----PPRLAIHTPSTAPSKFPPTPTDSQVSHSVYTONQ | 379 |
| QY | 666 | REDYAGOSHRNRNLGN--YPEKLSTRSHNAKNQPKSEKESENVQDGKGKHKHNR | 723 |
| Db | 380 | KQYTSKTHSHNSOQGTSMLEBDLISP-----SEDSSEQTP----- | 418 |
| QY | 724 | KRNEEGESESPLNELL-ETSRKCRGSSGIDETKDTLVLSRDAFTVRDEPMDA | 782 |
| Db | 419 | -----EKPPSSAPPASQSLPEFVVAASHSSASBESTSDS-----DSSSDS | 460 |
| QY | 783 | ESITFKSVSDRK-REKDKRYKSDTKRKSDGASATAKNDVLYKSGQOEKVDGDRKS | 841 |
| Db | 461 | ESSESSSDSENEPLETPAPEPEPTTNKQWQDNLMTKVQPAAPPEGRS-----TEP | 514 |
| QY | 842 | PRSEPLAKKAEKTKIDSKVPSSSSQKDEKYGTPRKHAHNSAADTROSPRTRRSKR | 901 |
| Db | 515 | PRRPEKSGSDSAT-----SQEHSSEKDPKPKSSAPRAPPEAPHP--GKR | 560 |
| QY | 902 | TVPRTSSQKQPYR---TRPRSLKRIWYLAREKNEBERKKKSDKDFESSKISKV | 957 |
| Db | 561 | SCQSPAQOEPPOQOYVGTQOPKKPK-----ASAAAGRTSLQGEFREGLLPYGR | 612 |
| QY | 958 | EGTEIVKPSFKRWEGVEKLETPPEKDIIASSTPPAKIKILNRETGKIGNAENASTTK | 1017 |
| Db | 613 | DQTS--KDKRKVTKG-----RPAALASN | 634 |
| QY | 1018 | EPSEKLSSTSKIKOEKVKKAKRYVAGSGSSSTLVDT-----STSSGSGPYAKS | 1070 |
| Db | 635 | EPKRAVPSPSEKKHKNSSLAPSKALSGPPADANVEDKTPHFLVLPILTESQGGPHSGS | 694 |
| QY | 1071 | EKTDTRRTYIKTMEYNNNDNTAPAEVDYIMIQVPOSKMDKDDFSESEEDVYTPPIQSV | 1130 |
| Db | 695 | GSRTSGROAVVVOEDSRKOR-----LPRLP-----RDTKLILSLDPT | 732 |
| QY | 1131 | GKPSIITKNVTTKPSATAKTESEQPEKLOKIPKASHLMQHELNNSKGSASSEGR | 1190 |

```

Db 733 PPGQSLMWKITL-----DLISRLIPQ-----PGKG---SROKR 762
QY 1191 AKDREHSGSEKONDPDKRKGAGQPDKESTVDRLSEQGHFKTLLSOSKKTETTSSEKHESYVRS 125
Db 763 AEDKOPRPAKKHSEKSSKSA-----SSSSKTLAKRRKGCEARDC 799
QY 1251 SNKQETPPRDKKVVYDSDYSSSKRRNDEGELAR-----RKDSPPGKRKSLSGOK---- 130
Db 800 DNKKIR--LEKEIKSQSSSSSSSKESKTKTPSRSSQSSKKEMLPPEPVSSSQPKAP 857
QY 1301 --SKLREERD-----PKGAESEKSN--SSPPDKK-----PHDK-APEYTKRP- 134
Db 858 ALKRRRRADTCGGDDPRKSASTSTSNHKDSSIPQRRVBEKSGRSSEHKGSSGDTANPF 917
QY 1342 -----CEETKPVDKNSGKEREKHAAEARNKESS 1370
Db 918 PVPSPLENGNSRKPQYKFDKQADLHMREAKKKMKOKA 955

RESULT 5
US-08-320-559-26
; Sequence 26, Application US/08320559
; Patent No. 5633135
;
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaan, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
TITLE OF INVENTION: All-1 Region
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5633135r1s
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/320,559
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/062,443
FILING DATE: 14 MAY 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/971,094
FILING DATE: 30-OCT-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/888,830
FILING DATE: 27-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/805,093
FILING DATE: 11-DEC-91
ATTORNEY/AGENT INFORMATION:
NAME: DeLuca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJD-0855
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1210 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-320-559-26

```

| | | | | |
|-----------------------|-------------------|--------------------|-------------|--------------|
| Query Match | 3.7%; | Score 276.5; | DB 1; | Length 1210; |
| Best Local Similarity | 18.3%; | Pred. No. 3.3e-10; | | |
| Matches 227; | Conservative 160; | Mismatches 428; | Indels 423; | Gaps 55; |

| | | | |
|----|------|--|------|
| QY | 231 | DIITFVSIYVSEKSDGFRSDOKL-LPAAALITSEHSKASSALITATLMEKGVGTSP | 289 |
| Dp | 66 | EWEKEFLSTKSTHRLDA-----SENRKLKPYPLPD--KSS-----IPSSF | 107 |
| QY | 290 | WNSIFVGOSLHGOLI--PTTGPYRIN--AARPGGRPWE-----HSNKLGIVLSPQ- | 339 |
| Dp | 108 | HTSV-----HQOSIHTPAGSLPVGNISHNPKMAQPRTEPMSPLAKSCG---PPDS | 156 |
| QY | 340 | -----QIRGGECSCTRSINRGHHSEBSORPOSPLATPCGVVYPPPLYPPLPHLLP | 395 |
| Dp | 157 | QHLTDRIGQGEFGSS-----HHKKGDRADG-----DHCAVY | 190 |
| QY | 396 | PGVPPQFSP--QPPSSOPPTAGSYVPPGFPAPAPANISTACSPGVTAHSN--TMPTT | 451 |
| Dp | 191 | DSAPERELSPILSPVSPV-----LSP-----IHSNQGLPFR | 224 |
| QY | 452 | QAPLRSREFRQONDKREKRFYSSSSYSRSSY--TDSQGLAOIH-----A | 499 |
| Dp | 225 | QG-----SSK--VHSSSNMSKOYCPAKSPKDLAVYVHDKETPODSLVA | 265 |
| QY | 500 | LTLSPSAHITDLDHHPHPEEAEARSAMIVNHPDLMJAHNRSRPPRYRSTRSRSP | 559 |
| Dp | 266 | PAQPS-----QTFPPRLSPKSVAMQOKP--TAYVRPD-----GGDAQP | 304 |
| QY | 560 | PEFGQSTKKNVREKEKEIEFKRYEV-----PPYDIKAYRSYDFKDPPE | 609 |
| Dp | 305 | SE-----SELKPLEDYRQOQFEKTDLKVPAKALTKLKMPSQSVQTSYNEHCVEILL | 360 |
| QY | 610 | KERYREWERKREYERYKUYAGAOPRANEDPSPRLPLYNIRN---SPETGR | 665 |
| Dp | 361 | KEMTHSW-----PPLAIHTPSTAEPSKEPPPTKDSQHVSVTQNG | 402 |
| QY | 666 | REDYAAQOSHNRNLGYN--YPEKLTSTRDSHNKADNKSEKSESENVPGDGKNKKKH | 723 |
| Dp | 403 | KOYDTSSKTHNSOOGTSSMLEDLQJLSD-----SEDSSEQOP----- | 441 |
| QY | 724 | KRNEKEKEESEFLNPELL-ETSRKRCGSSGIDETDTPFLPSKDATVYRDEPMDA | 782 |
| Dp | 442 | ---EKPPSSAPPSAPPOSILPEPYAAHSSASESTSDS-----DSSDS | 483 |
| QY | 783 | ESITFKSVSDMK-REMDKPVKSDTKRKSSDGSATKKNVNLPSKGPQEKVDGDRS | 841 |
| Dp | 484 | ESESSSSSENEPLETPAPEPEPTTNKQOLDWLTVKSQAPAPPGPS-----TPE | 537 |
| QY | 842 | PRSEPLLKAKEEATKTDIVYFSSSSOKDEKVTGTPRKAHKSADKTRROQPRTRSRK | 901 |
| Dp | 538 | PRRHPESKSSDSAT-----SQHSESKDOPPKSSKAPRAPEARHP---GKR | 583 |
| QY | 902 | TVRPTSSQSKOPVR-----TRPRSLKINLILAREKREKRRKKSVDKPRESSMKISKY | 957 |
| Dp | 584 | SCQSPAPQOEPPOQVTGCTOPKRPV-----ASARGSTRSLQGEHEPGLLYGSR | 635 |
| QY | 958 | EGTEIVRSPPRKKAEGVLEKERTPEKDIASSYTPPAKIKLIRETGKKIGNAENASTK | 101 |
| Dp | 636 | DQTS--KDKPVLTKG-----RPRAAAN | 657 |
| QY | 1018 | EPSEKLESTSKIOEKYKGAAKKAVGSESSSTLYDY-----STSSIGSPYKRS | 107 |
| Dp | 658 | EPKPAVPPSSSEKKKHKSSLPAPSAFSLGPPAPADNVEDRBEHFALVPLTESGPHSGS | 717 |
| QY | 1071 | EEKDITKRTVKTMEKYNNDTAPAEVYIIMIQPOSKMDKDFESEEBEDVKTQTOP | 113 |
| Dp | 718 | GSRTSGCRAAVVYGEDBKRD-----LPLPL-----RDKTLSPLRQT | 755 |
| QY | 1131 | GKPSIITKNYTKRSATAKTTEKSEQPEKLOKLPKASHMLQHELRSKSGASSEKGR | 1199 |
| Dp | 756 | PPPOSLWVTKT-----DLISRIPO-----PKG--SRQK | 785 |

```

0Y 1191 AKDRHSGSEKNDPNKSGAQPDKSESTVDRLSEQGHFKTLPSSQSKERTRTSEKSESVGCS 1250
Db 786 AEDKOPPAKGHSHSEKRS-----SDSSSKLAKRRKGAEHDC 822
0Y 1251 SNKFFTPGRDKKVVYDSNDYSSSKRRDERGELAR-----RKDSPPRCKESLISGQR---- 1300
Db 823 DNKRTR-LEKEIKSQSSSSSSSHKSESKTKPRSPRSSQSKEMLPPEPVSSSQKAPK 880
0Y 1301 --SKIREERDL----PKGAESKSN--SSPPRDK-----PHDK-APYETKRP- 1341
Db 881 ALKRSRRRDLTCGDOPPPKSSASTKSNHKKDSSIPKORRYEGKGSRRSSSEHKSGSDGTAMPF 940
0Y 1342 -----CEETKPVDCNSGKERKHAEAARNGKSS 1370
Db 941 PVPSLPNGNSPKRKPQYKFDXQADLHREAKMKOKA 978

RESULT 6
; Sequence 26, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities In the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESS: No. 6040140-ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10930
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/327,392
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/320,559
; FILING DATE: 11-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,443
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,094
; FILING DATE: 30-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/888,839
; FILING DATE: 27-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,093
; FILING DATE: 11-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229

```

```
? TELEFERENCE/DOCKER NUMBER: TJU-1262  
? TELECOMMUNICATION INFORMATION:  
? TELEPHONE: (215) 568-3100  
? TELEFAX: (215) 568-3439  
? INFORMATION FOR SEQ ID NO: 26:  
? SEQUENCE CHARACTERISTICS:  
? LENGTH: 1210 amino acids  
? type: amino acid  
? TOPOLOGY: linear  
? MOLECULE TYPE: protein  
? IS-Org:545-860D -26
```

| | | | | |
|-----------------------|-------------------|--------------------|-------------|--------------|
| Query Match | 3.7%; | Score 276.5; | DB 3, | Length 1210; |
| Best Local Similarity | 18.3%; | Pred. No. 3.3e-10; | | |
| Matches 227; | Conservative 160; | Mismatches 428; | Indels 423; | Gaps 55; |

| | | | |
|----|-----|--|------|
| QY | 231 | DITATVSIVSVSEKSDOPFRSDOKL-LPAAALTSEHSQASITATLAMEEKVGSTP | 289 |
| Db | 66 | EYKEFLSTKSHTHLDA-----SENRKLGKPYPLIPD--KGS-----IPSSF | 107 |
| QY | 290 | WNSIFVGOSLHGQL--PTTGPYRIN--AAREGGAPGME-----HSNKLGYLVSPQ- | 339 |
| Db | 108 | HTSV-----HHQSHTPASGLSVGNISHNMKMAOPRTPEMPSLHAKSCG---- | 156 |
| QY | 340 | -----QIRGSESCSRJNRRHHSERQORQOSLATECFYVYPPPLPYPPHTLPP | 395 |
| Db | 157 | QHLTQDRJGQCEFGSS-----HHKKGRRADG-----DHCAVY | 190 |
| QY | 396 | PGVPPOFSP--QEPSQOPPTAGSVPPGFPAPANISTAPCSPGVPTAHSN--TMPYT | 451 |
| Db | 191 | DSABERELSPILSPVPP-----LSP-----IHSNQTLPRT | 224 |
| QY | 452 | QAPLLSREEFYREONDKGRESKPEYSGSSYSRSTY--TSSQGLAHH-----A | 499 |
| Db | 225 | QG-----SSK--VHSSSNKSXGCGYPAKSPKDLAKVHDKENQDSTVA | 265 |
| QY | 500 | LTLSPSAHNTLDLHDPHPPEAEKASAMIVIMPDLMDLAHAKSSPPYRKRKSRSP | 559 |
| Db | 266 | PAOPPS-----QTFPPPLSPKSVAMQKP--TAYVRPD-----GQDQAP | 304 |
| QY | 560 | PEFRQSPPTKKNVREKEREYEFMRKYREP-----PRYDIKAYGRSVDPRPE | 609 |
| Db | 305 | SE-----SPELKPLPEDYRQOTFEKTDLAKYPAKAKLTKKMPQSVCOTYSNEYHCYEEL | 360 |
| QY | 610 | KERYREWKRYREMYEKYKGYAVAGAPRDSANREDFSPERLLPIRIN--SPTTGR | 665 |
| Db | 361 | KEMTHSW-----PPPLTAIHTPSAEPKPPFTKDSQHVSVTQNO | 402 |
| QY | 666 | REDYAAQOSHNNRNLGON--YPEKLTSTRDSHNAKNDPKSEKESSENVPGDGKGNKKHRR | 723 |
| Db | 403 | KOYDTSKTSHNSOQGSSMLEDLQJSD-----SEDSSESTP----- | 441 |
| QY | 724 | KRNRKEKEESESFLNELL-ETSRKRCGSSGIDETKTDTFLPSRDATPYDEPMDA | 782 |
| Db | 442 | -----EKPPSSAPPASAPQSLPEPYASHSSASESTSDS-----DSSSDS | 483 |
| QY | 783 | ESTFYSVSDOK-REKDKPKVMSDKTKRRSDGASATAKDNVLKPSKGQOEKYVDGREGS | 841 |
| Db | 484 | ESESSSSDSEENELFTPAPEPEPEPTTNKQLDNWLTKYSQAPAAPPEGPRS-----TEP | 537 |
| QY | 842 | PRSEPLTKAAKEEATKIDSVKSSSSQOKDKYGTGPKAHSTKAKXOTRQOSQORTRSR | 901 |
| Db | 538 | PRHPDEKGSDSAT-----SOHSESKODPPKSSSKAPVAPPEAPH-----GKR | 583 |
| QY | 902 | TVPTTSQKSOQPV-----TRPRSLRKINTINLIAREKNERKRKSYVDKFESSSMKISKY | 957 |
| Db | 584 | SCQSPSAQOEBPOQOTGTQOPKKPVK-----ASARGSRTSLQGEHEPELIFYGR | 635 |
| QY | 958 | EGTELYVKSPPRKMEGDEVKLERIPEKDKTASSTPAKTIKLNRREGKTIAGNAENASTK | 1011 |
| Db | 636 | DQTS--KQKPPVTKRG-----RPRAAAGN | 657 |

| | | | |
|----|------|--|------|
| QY | 1018 | EPSEKLESTSEKIQEKKVKKAKRRAKVGSSSSSTVDT-----SINSSGSPVRS | 1070 |
| Db | 658 | EPKPAVPSSSEKKHKKSSLPAPSAALSGPPADANDERTPEHFALVPLTESQGPBHGS | 717 |
| QY | 1071 | EKTDTRKTVTKTEEVNNNDTAPAEVDLIMIOVPOSKMDKDFESEEDVYTTPIOISV | 1130 |
| Db | 718 | GSRTSGCQAVVGEDSRKR-----LPLPL-----RDTLLSLPLDT | 755 |
| QY | 1131 | GKPSIIKNVTTKPSATAKYTEKESEQPEKLOKLPKASHELMHQHELNSSKGSASSEGR | 1190 |
| Db | 756 | PPPOSIMWKITL-----DLSTRIPQ-----PGK-----SRQK | 785 |
| QY | 1191 | AKDBEHSSSEKDNDRKRSQAQPKKESTVDRLSEQHFTLSQSKSEKRTSEKHEHSVGS | 1250 |
| Db | 786 | AEDKQPPAGKHHSEKSS-----SDSSSKLAKKKKGZHERDC | 822 |
| QY | 1251 | SNKDFTPGRKQVYDSDRYSSSKRRDERGELAR-----RKDSPRGESLSGOK--- | 1300 |
| Db | 823 | DNKKIR--LEKEIMSSQSSSSSSSHKESKTKRPRSPSSQSKMELPPVYSSSQKPAKP | 880 |
| QY | 1301 | --SKIRERDL-----PKGAEBSKSN-----SSPRDK-----PHDK-APETKRP- | 1341 |
| Db | 881 | ALKRSREADPCGODPPKSASTSNHKDSITKORVBEKGRSSSEHKGSSGDTANPF | 940 |
| QY | 1342 | -----CETKPVVDKNSGKEKEKHAARBGKSS | 1370 |
| Db | 941 | PVPSPILPNQNSKPGKPYVKFTDKQQAADLHMRPAKKMKQKA | 978 |

```

1      RESULT      7
2      PCT-US94-04496-26
3      ; Sequence 26, Application PC/TUS9404496
4      ; GENERAL INFORMATION:
5      ; APPLICANT: Croce, Carlo
6      ; APPLICANT: Cnaan, Eli
7      ; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
8      ; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
9      ; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1
10     ; NUMBER OF SEQUENCES: 86
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
13     ; ADDRESSEE: Norris
14     ; STREET: One Liberty Place, 46th floor
15     ; CITY: Philadelphia
16     ; STATE: Pennsylvania
17     ; COUNTRY: USA
18     ; ZIP: 19103
19     ; COMPUTER READABLE FORM:
20     ; MEDIUM TYPE: Floppy disk
21     ; COMPUTER: IBM PC compatible
22     ; OPERATING SYSTEM: PC-DOS/MS-DOS
23     ; SOFTWARE: Patent In Release #1.0, Version #1.25
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: PCT/US94/04496
26     ; FILING DATE:
27     ; CLASSIFICATION:
28     ; ATTORNEY/AGENT INFORMATION:
29     ; NAME: Deluca Esq., Mark
30     ; REGISTRATION NUMBER: 33,229
31     ; REFERENCE/DOCKET NUMBER: TJU-1242
32     ; TELECOMMUNICATION INFORMATION:
33     ; TELEPHONE: (215) 568-3100
34     ; TELEFAX: (215) 568-3439
35     ; INFORMATION FOR SEQ ID NO: 26:
36     ; SEQUENCE CHARACTERISTICS:
37     ; LENGTH: 1210 amino acids
38     ; TYPE: amino acid
39     ; TOPOLOGY: linear
40     ; MOLECULE TYPE: protein
41     ; PCT-US94-04496-26
42
43 Query Match      3.7% ; Score 276.5 ; DB 5 ; Length 1210 ;

```

Best Local Similarity 18.3%; Pred. No. 3.3e-10;
Matches 227; Conservative 160; Mismatches 428; Indels 423; Gaps 55;

```

QY 231 DITATVSVSHSEKSDPFSDDKL-LPAALTEHSEKSGASSAIALAMEKVPCTSP 289
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 EVELSTKSTHRLDA-----SENRLGPKYPLIPD--KSSS-----IPSSP 107
QY 290 WNSIFVQSLHGLI--PTTGPYRIN--AAPGGRRPGE-----HSNKLGYLVSPQ- 339
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 HTSV-----HQSIITPASPGLSVGNISHPKMAQPRIDPMSLAKSGC-----PPDS 156
QY 340 ----QIRGERSCYRINRGHSEKQRTQSPSLPATPCFVVPVPPPLPPTPLP 395
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 QHUTQDRIGQEGEGSS-----HHKKGDRRADG-----DHCAAVT 190
QY 396 PGVPPQESP--QEPSSQPTAGSVPPPGPPAPANISTACSPGVPTAHSN--TMAPTT 451
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 DSAPERLSPILSPVPP-----LSP-----IHSNQTLPRT 224
QY 452 QAPLLSREFFYREONDKGRSKPPYSGSYSRSSY--TDSQGLAQHIIH-----A 499
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 QG-----SSK--VHGSSNNKGYCPAKSPKDLAVKHDKETQDSLVA 265
QY 500 LITSPSAHLTLDLHPHPPEEEAARSAMIVHMPDLMDIAHARSPPYRRTYRSRSP 559
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 PAOPPS-----OTFPPPSLPKSKVAMQOKP--TAYVRPMD-----GDOQAP 304
QY 560 PEPFGOSPTARNVPREKREYFNRYREV-----PPYDIKAYGSRVDFRPE 609
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 305 SE-----SPELKPDPEDYRQOTFEKTDLKVPAKAKLTKLAKMSQVSEOTYSNEVCHVEIL 360
QY 610 KERYREMYREMYEKYKGYAVGAQPPRPSANREDESPERLPLNIN--SPFTGR 665
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 KEMTHSW-----PPPLTAIHTPSTAEPKPFPTKDSQHVSVYQNO 402
QY 666 REDYAAQSHRNRLGCGN--YPEKLTSTRDSHNAKDNKSKKESENPVGGDKGKKHKKR 723
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 KQYDTSSKTHSNSQGTSMLEDLQDLS-----SEDSDEQTP-----441
QY 724 KRREKEGESESPINPELL-ETSRKCRGSSGIDETDTLFLVPSRDAATPVHDEPMDA 782
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 ----EKPPSSAPPASQSLPEPYASAHSSAESESTDS-----DSSSDS 483
QY 783 ESTFEKSVSKDK-REKDKPKVSKDKTKRKSDGSAATAKKNVLPKSGPOEKVDREKS 841
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 484 ESESSSSDSENEPLETPAPEPEPTTNKQDLMWLTKVSOAPAPPGSPRS-----TEP 537
QY 842 PRSEPPLAKKKEATKIDSVKPPSSSOKDKVGTGTPRKAHKSAAKDRROQSPTRRSKR 901
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 538 PRHPESKSGSDSAT-----SOEHSESKDPKPKSSSKAPRAPEAPH--GKR 583
QY 902 TVPTSSQKQPVYR--TRPRSLRKINYLAREKNEREKRSKVDKDESSSMKISKV 957
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 584 SCQKSPAQOEPPOQOTGTQPKPKPV-----ASARAGSRKSLGERRPGLLPGRS 635
QY 958 EGTETVPSPKRMKEGVEKLETPERKOKIASSTTPAKKIKLNETGKKIGMAENASTTK 1017
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 636 DQNS--KDKPKVTKG-----RPRRAASN 657
QY 1018 EPEKLESTSSKIKQEVKKGAKKAVAGSEGSSTLYDYT-----STSTGSSPYRKS 1070
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 EPRPAVPPSEKKHKSSLPAPSKALSGPREPAKDNVEDRTPENHALVPLESQPSPHGS 717
QY 1071 EEKJDTKRTVIKMEENNNNTAPAEVIMIOVQSKWPKDDESEDEVDKTTQPTOSV 1130
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 718 GSRTSGCROAVVOEDSRKR-----LPLPL-----RDKLILSPILDT 755
QY 1131 GKPSIINKNTTPSATAKYTEKESEQPERKLOKLPKEASHLMQHELRSKSGASSEKGR 1190
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 756 PPQSLWVKITL-----DLLSRIPQ-----PGKG--SRQK 785
QY 1191 AKDREHSGSKNDPKKSGAQDPKESTVDRLSEQGHFKTILSGSKKTRTSEKHESVRGS 1250
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 786 AEDKOPPAKRRKHSSEKRS-----SDSSSKLAKRRKGEAERDC 822
QY 1251 SNKDFTEGDKKVDYDSRSSKRRDERGLAR-----RKDSPPKGESLSGOK---- 1300
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 823 DNKKIR--LEKEIRSGSSSSSSKHSKSKTPSPSSQSKKMLPPPPVSSSQKAPK 880
QY 1301 --SKUREERDL--PKGAESKRSN--SSPPDK-----PHDK-APYETKRP- 1341
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 881 ALKRSREADTCGDDPPKASSTRKSNHKDSSIPKORVEKGSRSSSEHKSSGDTANPF 940
QY 1342 -----CEETPKVDKNGKEREKHAAPARNKESS 1370
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 941 PVPSLPNGNSKPKQYKFDKQADLHMRKMKKQKA 978

```

```

RESULT 8
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THILVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
; US-07-741-940-7

```

Query Match 3.7%; Score 274; DB 1; Length 2842;
Best Local Similarity 18.3%; Pred. No. 1.5e-09;
Matches 282; Conservative 188; Mismatches 612; Indels 446; Gaps 61;

| | | | |
|----|------|--|------|
| 0Y | 173 | FRVYVSPCSDPTKAGCSCSPGSLSLPAPSTLSLNSG--SLAPVYSGNPPSSAPAVP | 230 |
| Db | 1194 | FSEKSSSSSGSKTEHMSSESTTS---TPSSAKRQNDLHPSAKSSGQOKA---- | 1245 |
| 0Y | 231 | DITATVTSIVHSEKS-----DGPF---RDSDNKLPLPAA-----LTSEHSGASSIA | 274 |
| Db | 1246 | ---ATCVYSSINDQETIGTVCEVDPTICIFSNCSLSSLAEDIEICNQTTQDASANTLQ | 1302 |
| 0Y | 275 | ITALMEEG-----VPTGTSP-----WNSITVGSLLHGQILPTTG---PVR | 312 |
| Db | 1303 | IAETKEKIGTRESADPYSEVPAAVSPQHPRTKSRKSLQSSLSSESAKHAKVESGSAKPSK | 1362 |
| 0Y | 313 | INAPRGGRGPRGHEHNSKLTGLVPSPOQOILRGERSCYRSINGRHHSEKSOITGSPSLPA | 372 |
| Db | 1363 | SGAOTP-----KSPBEHVQETPLMFSCTSCTVSSLSLDESNSIASSVQS | 1406 |
| 0Y | 373 | TPCFVYV-----PPPLYPPPPHTPLRLPGVPPOFSP---QFPSSQPTAGYSVP | 420 |
| Db | 1407 | EPSCGMSVGLISPDLLDPSGOTMPPSRSKTP--PPQTAQTKREVPRKNKAPTAKRES | 1464 |
| 0Y | 421 | PRG-----PPRPA--NISTACFSGVYTAISNTMPTQOAPLSREFFYR-- | 463 |
| Db | 1465 | GPKOAAVAAVORYOVLPLDADTLHLHATBSTPDRGFCSSSLALSLSIDEFIQKDVLEIM | 1524 |
| 0Y | 464 | ---BQNDKGRSKPRYSGSSYSRSTYDSSOGIAQIHALTLPSPAHTLIDLHNPHP | 520 |
| Db | 1525 | PRVQENDNGNTE-----SQPKESMNDQEKAEKTI DS-----EKLLDSDSDDD | 1570 |
| 0Y | 521 | EBEARSAMIVHMDLMDIAHARSRSPRYKRSRSRSPFEER---GOSPTKRNVPREK | 577 |
| Db | 1571 | IEI-LEECILISAMP-----TKSSRKAKPAQATASKLPPAPARRKPOLPYKLLPSQNR | 1622 |
| 0Y | 578 | EREYFNRYREVPREPYDIKAYV---GGSVDRPDQFEKERREMERKYYREMYEUYGYAVG | 634 |
| Db | 1623 | LQP---QKHVSTPEDDMPRYCYEGRPINSTATSISLDT-----IESPPNELAAG | 1671 |
| 0Y | 635 | AQPPRSANREDPSEERLLPLINRNSPFTTRGRREDVYAAGSHRNRLGNYPREKLSRSH | 694 |
| Db | 1672 | EGVGGAQSGEFEKROTI P-----TEGRSTDEAQ-----GTVSSVTIPELDN | 1715 |
| 0Y | 695 | NAKDN-----PKSKEKE-----SENVPG---DGKGNKHKKHR | 723 |
| Db | 1716 | KAEBOILLAECSINAMPKGKSHKPRFYKIKIMDVQOASASSAPKKNOLDKKKKPTSPV | 1775 |
| 0Y | 724 | K-----RNEKEGGESESFLINPELLTETSRKCRSSGIDETKIDTTLFVLRSDA--- | 772 |
| Db | 1776 | KPIQNTNEYRRRYAKNADSKNNLNLEBEVPSDNKDSKKOYLKNNNSDFNDKLPNNDRRG | 1835 |
| 0Y | 773 | -----TPVDEPM-----DAESITFKSVSDKDKREKDKPRKVSYDKTKRKSDDGA | 816 |
| Db | 1836 | SPAFDSPPHHYPIGTFYCFPSRNDSSLSLDF---DDDDVDLSREKALRKKEKKESEA | 1891 |
| 0Y | 817 | TAKDNVLPKPSGQOEVDVDREKSPRSBPPLKAKAEATKIDSYK--PSSSSQODEK-- | 872 |
| Db | 1892 | KVTSITELTSNQSANKTQALIAQPINRKOPAKITLOKSTFPQSSKDIPLDGAATDEKIQ | 1931 |
| 0Y | 873 | ---VTGTP-----RAHASKSANDTR--ROSQPTRRSKRTVPTS | 907 |
| Db | 1952 | NFAIENTPVCFSHNSLSSLSLSDIDQENNNKENEPKETEPDPSQGBPMQASGVAPKSF | 2011 |
| 0Y | 908 | SOKSQPVRTRRRLKRIYN---LI-----AREKNREK----- | 938 |
| Db | 2012 | HVEDTPVCFSSNSSLSSLSIDSEDDLQECISAMPKKKPKRLKGDNEKHSPRNMGIL | 2071 |
| 0Y | 939 | -----RKSVDK | 945 |
| Db | 2072 | GEDTLTLDKIQRDBSHGLSPSENPDMKAIQEGANSIVSSLHQAAAAACLSRQASDS | 2131 |
| 0Y | 946 | DE-----ESSSMKISKYEGEIVYKSPPKKMGVDEKLETRPEREKTI | 987 |
| Db | 2132 | DSILSKSGISLSPFHLTPQOEKEPPTSKGRRLLKPKBEKSTLE--TKKISESKGKIR- | 2188 |

| | | | |
|----|------|--|------|
| Qy | 988 | ASSTPPAKKILINNETGTICGNENMSTTPEPEKLESTSSKIKOEKVGAKRKVASE | 1047 |
| Db | 2189 | -----GGKKYKSLITICKVANSSEISGOMKOP--LQAMPSTIS-----RGRTMIHPGVR | 2236 |
| Qy | 1048 | GSSTLVDYVSTSTSGGSPVR-----KSEKTDTRVITKMEEYNNDTAPADVIYIMI- | 1102 |
| Db | 2237 | NSSSS-----TSPVSKGPPILKTSPASKSPSPGQATATSPRGAKPVYSKSELSPVARQTSIG | 2292 |
| Qy | 1103 | ---QVPOSKMDKDFEEDVDVTTQPIOSVGKPS-----SITKN | 1139 |
| Db | 2293 | GSSKAPRSRSGSRSTPBPAPQPLSRDIPSGNNISIPGRNGISBPNNKLSQLPRTTSPST | 2352 |
| Qy | 1140 | VTPKPSATK--VTE-----KESBOPEKLOKTRKEASHHELOHELSSKGSASS | 1186 |
| Db | 2353 | ASTSSSGSKMSTYSPERGMSQNLKQGLSKSNASSIPRSSASAKLGNOMNNGAKMK | 2412 |
| Qy | 1187 | EKRPAMDREHSSSEKDNPPK---RKSGAQDPKESTYDR--LSEQGHFKITLSOSK--- | 1236 |
| Db | 2413 | VELRMSSTKSSGSEDSRSERPVLVQOSTFIKAPPTLRRLLEASSTESLSPESSRAS | 2472 |
| Qy | 1237 | ETPTPSEK-----HEVYRSGSNKQFTPGCRDKKYDVOSPDYSSSKRRDERGE | 1281 |
| Db | 2473 | PTRSQAQTPULSPPLPMSLSTHSSVOAGGKRLPLNLSPTIEIN--DORPAKRND----- | 2526 |
| Qy | 1282 | LAR-RKDSPPRGKESLGC--OKSKLREERDLPKKGAEKSSKSNSPPRDCKPMDHAKPYET | 1338 |
| Db | 2527 | IARSHSESPRLPIINFGJTWKREHKSUSSLPRVSTWMRRGTSSSSITLSASSSESE----- | 2581 |
| Qy | 1339 | KRPCEETKPVUDKNGKGREKHAARENG | 1366 |
| Db | 2582 | KAKEDEKHNVISIGTQSKOENOVASAKG | 2609 |

RESULT 9
 US-08-289-548A-7
 : Sequence 7, Application US/08289548A
 : Patent No. 5648212
 :
 : GENERAL INFORMATION:
 : APPLICANT: ALBERTSEN, HANS
 : APPLICANT: ANAND, RAKESH
 : APPLICANT: CARLSON, MARY
 : APPLICANT: GRODEN, JOANNA
 : APPLICANT: HEDGE, PHILIP J.
 : APPLICANT: JOSLYN, GEORGE
 : APPLICANT: KINZLER, KENNETH
 : APPLICANT: MARKHAM, ALEXANDER F.
 : APPLICANT: NAKAMURA, YUSUKE
 : APPLICANT: THILYERIS, ANDREW
 : TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 : TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 : NUMBER OF SEQUENCES: 102
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Banner & Allegretti, LTD
 : STREET: 1001 G Street, NW
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20001-4598
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/289,548A
 : FILING DATE: 12-AUG-1994
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Kagan, Sarah A.
 : REGISTRATION NUMBER: 32,141
 : REFERENCE/DOCKET NUMBER: 1107.46943
 :
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2842 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: APC
 US-03-289-548A-7

Query Match 3.7%; Score 274; DB 1; Length 2842;
 Best Local Similarity 18.5%; Pred. No. 1.5e-09;
 Matches 282; Conservative 188; Mismatches 612; Indels 446; Gaps 61;

173 FRYTSPYPTCDTKAGSCSDGTLRLPAPISISLTNO--SLAPYVSGNPPSAPAPVP 230
 1194 FFSKSSSGSGSKTBMSSSENTS--TPSNAKRQQLHPSSAQSHSGGQPKA----- 1245
 221 DITATVSISSVSEKS-----DGF--RSDNKLPLAA-----LTSEHSGASSIA 274
 1246 --ATCKVSSINQETIQTYCEDTPICFSCSSLSLSSADEIGCNQTOGEADANTLQ 1302
 275 ITALMEEG-----VGTSP-----WNSIFVGOSLHGOLIPPTG--PVR 312
 1303 IAEIKETIGTSAEDPVSEVAVASHPTKSSRLQSSLSLSSANAKHAKVESSGAKSK 1362
 313 INAAPGGGPRGWEHNSKGLVSPQOIRGERSCYRSINGRHNSERSQTOGSPSLPA 372
 1363 SGAQTP-----KSPENHYQETPLMFSCRYSVSLDSFESRSIASVQS 1406
 373 TPCEFPV-----PPPLXPPPHPLPLPVGYPPOFSP--QFSSQPPPTAGIYVP 420
 1407 EPCSGMVGSIISPSDLPSPQOTMPPSRSKTPP--PPQOTQTKREVAKNAPTKEKES 1464
 421 PRG-----FPAPA--NISTACSPGVPTAHNTMPTQAPLLSREEFYR-- 463
 1465 GPKQAAVAVAVQVQVLPDALTLHLFATESPPDGFSCSSLSLSSLDPEFLOKVELIM 1524
 464 --BONDGRESKPPYSGSSYSRSSYDSSQLAOHIALTLSPSAHTLDDLHDHPRP 520
 1525 PPVQENDNGNETE-----SEQPKESNENQKEAEKTIDS-----EKDLDDSDDD 1570
 521 EEARSRAMLYHMDLMDIAHAKSRPPYRRYRSKRSRSPPEFR--GOSPTRKRVPREEK 577
 1571 IEI-LEECIIISAMP-----TKSSRAKAKPAOTASKLPPVARKPSOLPYKLLPSQNR 1622
 578 EREYFNRYREVPPIYDIKAVY--GRSYVDFDPPEKERYREMERKYRMYEKKYVGYAVG 634
 1623 LQP--QKIVSTTPGDDMRYVCVESTPIINFSTATSLSLT-----ISSPNELAAG 1671
 635 AQPSPANREDFSPERLLPLNIRNSPTLRGRREDYAAQOSHNRNMLGNYPEKLTSDSH 694
 1672 EGVGGAGSGEFEKRDTP-----TEGSTDEAOG-----GKTSSVTIPELDON 1715
 695 NAKDN-----PKSKEK-----SENVPG--DGKGNKHKHR 723
 1716 KAEGBDILAEICINSAMPKSHKPPRVKKINDVOOASASSAPKNOLDGKKKKPTSPV 1775
 724 K-----RNEEKGESSEFLNPELLETSSRKCGSSGIDETKDTLFLVPSRDA--- 772
 1776 KPIONTEYTRVRKKNADSKNNLNAERVSNDKSKKONLNNKSKENDKLPNNEDRYRG 1835
 773 -----TPVDEPW-----DAESTFYSVDKDKREKDKPKVSDTKRKSDDSA 816
 1836 SFAPDSPHHYPIEGTPIYCFGRNDSLSLD-----DDDDVDLSREKALKRAKKEKESEA 1891
 817 TAKDNVILPKSGPOEKYVDGDKREKSPNSEPPLKAKKEATRTIDSVK--PSSSSQKDEK-- 872

1892 KYTSHTELTLSNOQSANKTOAIKOPINRGOPKPILOKOSTEPPOSSKIDPDGATDEKQ 1951
 873 ---VTGTP-----RKASKSAKOTR---ROGOPTRRSKRTVPKTS 907
 1952 NFALNTPVCSHNSLSLSDIDOEENNNKENEPIKETEPPDSOGEPKPAASGAPASFS 2011
 908 SQSQPVTRRRPRSLRKINY-----LI-----AREKNEREK----- 938
 2012 HVEDTPVCFSRNSSLSSLSIDSEDDLQECISAMPKKKKKPSRLKGDNEKHSPRMGGIL 2071
 939 -----RKASDK 945
 2072 GEDTLDKIDQRPDSEHGLSPDSENFDMKAIQEGANSIVSLHQAAAAACLSRQSSDS 2131
 946 DF-----ESSMKISKVGTETIVKPSPKRMGDVEKLETPPEKDI 987
 2132 DSILSKGSLSGSPFHILTPQOEEKPFISNKGPRILKRGKSTLE--TKKIESEKGLK- 2188
 988 ASSTTPAKKIKLNRETGKIGNAENASTTKPESEKLESTSSKIKQEKVYKAKRRVAGSE 1047
 2189 ---GKKVYKSLITGKVRNSRSEISGOMKP--LQANMPSIS---RGRTHIHIPGVR 2236
 1048 GSSSTLVDTYSTSTSGSPVR-----KSEKTDYKRTVIKTMEEYNNONTAPAEVYIMI- 1102
 2237 NSSSS---TSPVSKKGPPLTPASKSPSEOTATTPSPGAKPSYKSELSPVAROTSQIG 2292
 1103 ---GVPOGKWDKDFESEEDVKTQPIQSVGKPS-----SIITKN 1139
 2293 GSKAKPSRSGSRDSTPPSPAPQPLSRPIQSPGRNISIPGRNGISPPNKLSQLPRSSPST 2352
 1140 VTKPSATAK--YIE-----KESEDEKLOKLPEKASHLMOHELRSKGSASS 1186
 2353 ASTKSSGSKMSYSPGROMSQONLTQTLGSKNASSIPRESASKGLNOMNNGANKK 2412
 1187 -EKGRAKREHSGSEKMDPK-----RKSQAOPDKESTYDR--LSPOGHFKLLSOSK--- 1236
 2413 VELSRMSTKSSGSEDSRSEPVLRQSTFIKEADSPYLRKRLKLESASFEESLSPSRAS 2472
 1237 ETRTSEK-----HESVRGSSNNKDFTPGRDKVYDYSRDYSSSKRRDEGE 1281
 2473 PTRSGAQTPIVLSPLPDMSLSTHSYQAGGWRKLPNLSPILEYN--DGRPAKRKD----- 2526
 1282 LAR-RKDSPPRGKESLSC--OKSKLREERDLPKKAESKSSSPPRDKKPHDKAPYET 1338
 2527 IARSHSESRLPIRISGTWKRREHSHSSSLPRVSTWRRGTSSSSILSASSESE----- 2581
 1339 KRCEETKPVDNKSGEREKHALEARNG 1366
 2582 KAKSEDEKHVNSIGTKOSKENQVSAGK 2609

RESULT 10
 US-08-452-654-7
 Sequence 7, Application US/08452654
 Patent No. 5691454
 GENERAL INFORMATION:
 APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner, Birch, Mckie & Beckett
 STREET: 1001 G Street, NW
 CITY: Washington

STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-452-654-7

Query Match 3.7%; Score 274; DB 1; Length 2842;
Best Local Similarity 18.5%; Pred. No. 1.5e-09;
Matches 282; Conservative 188; Mismatches 612; Indels 446; Gaps 61;

QY 173 FRYVSPSCSDTKTAGSCSDSGTSLRLPAPSTISLTSNQ--SSLAPVYSGNPSAPAPVP 230
DB 1194 FFSKSSSSGSSSKTEHMSSESENTS---TPSSNAKRQNLAPSSAQSSGQPKA----- 1245
QY 231 DITAVSVISVSEKS-----DCPF---ROSDNKLPLPAA-----LTSESKKASSIA 274
DB 1246 ---ATCKVSSINQETIQTVCVEDTPICFSCSSLSLSAEDICGNTTOGADSANLQ 1302
QY 275 ITALMEKG-----VPGTSP-----WNSIFVGSLHGLIPTTG--PYR 312
DB 1303 IAEIEKIGTSAEDPVSEVAVSHPRTKSSRLGGSSLSLSAHHKAVESSGAKSPK 1362
QY 313 INAAPGGGPRGWEHNSKGLVLPPOQIRGERSCYRSINGRHHSERQTOGSPSLPA 372
DB 1363 SGAQRP-----KSPPEHYVQETPLMFSSRCTSVSSLDSPESHSSIASVYS 1406
QY 373 TPCFPVY-----PPPLYPPLPHLPLPPVPPQFSP---QFESSOPTAGYSVP 420
DB 1407 EPCSGWGIISPSDLPSPGQTMPPSRSKTTP--PPQTAOTKREVPPKNKAPTAEKRS 1464
QY 421 PPG-----PPAPA--NISTACFSPGVPTAHSNMPPTQAPLRFEEFYR-- 463
DB 1465 GPKQAANAAYOVOLVLDADTLHMFATESPTDGFSSSSLSALSIDEFIOKVDLRLM 1524
QY 464 ---EONDGRESKFPYSGSSYSRSSYDSSOGLAQHIALTLSPSAHTLIDLHDPHP 520
DB 1525 PVOENDNGNENE-----SEQPKESNENQEKAEKTIDS-----EKDLDDSDDD 1570
QY 521 EEAERASMIYVPMILMIAHARSPPYRRKRSRSRSPERR---GQSPTKRVNPREK 577
DB 1571 IET-LEECTISAMP-----TKSSRAKAKPAQTASKLPPVAPKPSQLPVYKLLPSQNR 1622
QY 578 EREYFNRYREVPPIYDIKAYV--GRSYDFRDPFEKERYREWERKYREMYEKYKGYAVG 634

DB 1623 LQP---QKHVSFTPGDDMPRVYCEVETPIINFSTATSLDUL-----IESPPNELAAG 1671
QY 635 AQPSPANKREDESPERLLPLNIRNSPFTGRREDYAAQOSHRRNMLGNGYPEKLTSDRSH 694
DB 1672 EGVRCGAQSGEPEKRDITP-----TEGRSTDEAOG-----GKTSSVTIPELDN 1715
QY 695 NAKDN-----PKSKEKE-----SENVPG-----DGGNKHKKHR 723
DB 1716 KAEEDDILAEICINSAMPKQSHKPPRVKIMQVOQASASSAPKNDLQDKKKRPTSPV 1775
QY 724 K-----RRNEEGEESFELNPELLETSSKCGSSGIDTDTLTVLPSRDA---- 772
DB 1776 KIPONTETRTVVRKNADSKNNLMAEVPYSDMKDKQVULKNNSDPNDKLPNNEDRVAG 1835
QY 773 -----TPVRDEPM-----DAESTFFKVSDDKREKDKPKVYSDKTKRKSDDGA 816
DB 1836 SFAPDSPHHYPTIEGTPYCFSRNDSLSLD-----DDDVLDLSREKAEKRAKENKESPA 1891
QY 817 TAKKNVILKPSKGPQEKVDGDEKSPRSEPLKKAKEATKIDSVK--PSSSQDEK-- 872
DB 1892 KVTSHTELTSMQASANKQALAKOPINRGQPKPILQKOSTPQSSKIDIPDGAATDEKQ 1951
QY 873 ---VTGTP-----RKAHKSADTR--ROSOPTRRSKRTVPKTS 907
DB 1952 NFAIENTPVCFSHNSLSLSDIDQENNNKENEPKETEPPPSQGPSPKQAGYAPKSF 2011
QY 908 SOKSQPVTRRRPSLRKINY-----LI-----AREKNERK----- 938
DB 2012 HVEDTPVCFSRNSLSLSIDSEDLQECISSAMPKKKKPRLKGDNEKHSPPRMNGITL 2071
QY 939 -----RKSVDK 945
DB 2072 GEDLTLDLKDIORPDSEHGLSPDSENFPMKALQEGANSTVSSLHOAAAAACLSRDASSS 2131
QY 946 DF-----ESSMKISKVEGETEIVKPSPKRMGADVEKLETRPEKDKI 987
DB 2132 DSILKSGISLISGPFHLTPDDEKPFPSNKGPRILKPEKLTLE--TKKISESKGIR- 2188
QY 988 ASSTTPAKKIKINRETGKKIGNAENASTTKPESEKLESTSSKIKOEKVKAKRRVAGSE 1047
DB 2189 -----GKKVYKSLITGKVRNSSEISGOMKP---LOAMPSPIS---RGRTMHIIPGV 2236
QY 1048 GSSSLVDVTSSSTSGGSPVR--KSEKTDTKFTYIKTMEYINNDNAPADVIIM- 1102
DB 2237 NSSSS---TSPVSKKGPPLKTPASKSPSEGTATTPRGAKPSVKSSELSPVARQTSQIG 2292
QY 1103 ---QVPQSKMDKDESEEDVKTTOPIQSVGKPS-----SIIRN 1139
DB 2293 GSSKAPSRGSDSTPSRAQOPLSRPIOSPGRNLSIPGRNLSIPPNKLSQLPRTSSPST 2352
QY 1140 VTKPSATAK--YTE-----KESEQPEKLOKLPKEASHLMOHELRSKSGSASS 1186
DB 2353 ASTKSSGCKMAYTSPGRMSOONLTKGTGLSKNNSIPRESASAKGLNONGNANK 2412
QY 1187 -EKGRAKDREHSGSEKDNDK--RKSQAOPDKESTYDR--LSQGHKKTLSQSK-- 1236
DB 2413 VELSRASSTKSSGSDSRSEVVLVROSTFIEAAPSPLRRKLESASSESLSPPSRPAS 2472
QY 1237 ETRTSEK-----HEVRSKSNKDFPPGRDKVDDYSDRVSSSKRDERGE 1281
DB 2473 PTRSQAQTPVLSPLPDMKLSHSSVQAGWKRLPNLSPLTIEYN--DORPAKRHD---- 2526
QY 1282 IAR-RKDSPPRGKESLIG--QKSKLREERDLPPKKAESKSSNSPPRDKKPHDKAPYET 1338
DB 2527 IARSHSESSRLPIRNSGTAKREHSHSSSLPRVSTWRTGTSSSILSASSSESE----- 2581
QY 1339 KRCEETKPVDKNSGKEREKHAABARNG 1366
DB 2582 KAKSEDEKHVANSISGTQSKQENQVSAKG 2609

RESULT 11


```

Oy 1187 -EKRAADREHSSGSEKNDPDK-----KKSQAQDPKESTYDR--LSQGHFKITLSQSK---1236
Db 2414 VELSRMSSTSSGSSDSRSPALVYHQSFIFKEAPSPTLRRRLKLESASPESTLSPSSRPAS 2473
Oy 1237 ETPRSEK-----HESVRGSSNKDFFPGGRDKYVDYSPRYSSSSKRRDERGE 1281
Db 2474 PTRQQAQTPVLPSLPRLPMLSLSTSSVQAGGWRKLPNLSPTIEYN--DGRPAKRND-----2527
Oy 1282 LAR-RKDSPPRGKESLSG--QKSKLREERDLPKKAESKKSNSPPRDKKPHDKAPYET 1338
Db 2528 IARSHSESPRLPRLNRGCTWRKRESHKSSSLPRLVSTWRTGSSSSILSSASSESE-----2582
Oy 1339 KRCEETKPYDKNSGKREKHAHAELANG 1366
Db 2583 KAKSEDEKHVNSISGTOKSKENCVSAGK 2610

RESULT 12
Sequence 7, Application US/08452655B
Patent No. 5783666
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES

```

[illegible]

```

Db 2073 GEDLTLDLNDIQRPDSEHGLSPDSENDMKAIOEGANSIVSLHQAANAACLSRQASDSDS 2132
QY 946 DF-----ESSSMKISKVEGEIYKPSKRMKEGVLEKPTPEKDKI 987
Db 2133 DLSLSKSGISLSPHLPDOEKPEPTSNKGPRILKPGKSTLE--TKKIESESOKIK- 2189
QY 988 ASSTTPAKKIKLRENGKIKGNENASTKEPSEKLESTSSKIKOEKVKAKRKRVAGSE 1047
Db 2190 -----GKKKVKKSLITGKVNSEISQOMKQP---LOANNPIS---KRTMIHIGVR 2237
QY 1048 GSSSTLVDTSTSTGSGPVYR---KSEKTDKRTVYIKTMEEYNNNDNTPADVYIIM- 1102
Db 2238 NSSSS---TSPVSKKGPPLKTPASKSPSEGGATTSPPGAKPSVKSLSLPAVROTQIG 2293
QY 1103 ---QVQSKKDKDFESEEDVKTOTPIQSVGRPS-----SITKN 1139
Db 2294 GSSKAPSRSGSRDSTPSRPAQPLSRPIQSPGRNSISPGRNGISPPNKLSQLPRTSPST 2353
QY 1140 VTKRPSATAK--YTE-----KESQPKLQKLPKEASHLMOHELRSKSGSASS 1186
Db 2354 ASKSGSGKAMSTTSPGRQMSQONLTKOTGLSKNASSIPRSEASKGLNMMNGANKK 2413
QY 1187 -EKGRADREHSGSEKDNPKD---RKSGAPDKESTVDR--LSEOGHFTLSQSSK-- 1236
Db 2414 VELSRMSSTKSGSESDRSERPVLVROSTFIKEAPSPTLRKLLEASAFESLSPSRPAS 2473
QY 1237 ETTTSEK-----HESVAGSSNKQFTPGRDKKVYDSDYSSSKRRDERGE 1281
Db 2474 PRTSQOQTPVLSPLDPMLSLTHSSVOAGKRLPPLSLPTIETYN--DGPARRHD---- 2527
QY 1282 LAR-RKDSPPRGKESLSG--QKSKLREERDLPKKAESKSNSSPPRDKRPHHKAPYET 1338
Db 2528 IASHSHSPRLILNSGTAKREHSHSSSLPVSSTWRRIRGSSSSILSASESEF----- 2582
QY 1339 KRPCETKPVKNKSGKEREKHAABARNG 1366
Db 2583 KAKSEDEKHVNSISGTQSKENQVSAKG 2610

```

```

RESULT 13
US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MART
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THILVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amlng acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-450-582-2

```

```

Query Match 3.7%; Score 274; DB 3; Length 2843;
Best local Similarity 18.5%; Pred. No. 1,5e-09;
Matches 282; Conservative 188; Mismatches 612; Indels 446; Gaps 61;

```

```

QY 173 FRYVSPSCSDRTTASGSDGTLRLPAPISLTSNQ--SSLAPVSGNPSAPAPVP 230
Db 1195 FSPKSSSGSSKTEHMSSESSNTS--TPSSNAKRONQLHPSSAOSRGOPKA----- 1246
QY 231 DITATVSIYSHSEK-----DGPFF--RDSNKLILPAA-----LTSEHSGKASSTA 274
Db 1247 ---ATCKVSSINQETIQTVCEDTPICFSCSSLSLSSAEDEIGCQNTQEDSANTLQ 1303
QY 275 ITALMEKQ-----VPGTSP-----WNSIFVGSLHGLIPTTG---PVR 312
Db 1304 IAIKEIKIGTSADEPVEPVPANVQHPRTKSSRLQSSLSLSEBARHKAIVFSGAGASPSK 1363
QY 313 INARPGGGRPWGSHSNKGLVYSPQOIRGRGSCYRSINGRHHSNSQRTQSPSLPA 372
Db 1364 SGAQTP-----KSPPEHYVQETPLMFSSRCTSVSSLDSEFSRSIASSVOS 1407
QY 373 TPCGFVPY-----PPPELPPPPHTLPPGVPPPOSP---QPPSQPTAGSVY 420
Db 1408 EPCSGMVGSLTSPSDLPSDFGOTMPPSRKTPP--PPQTAOTKREVPKKAFTAEKRES 1465
QY 421 PPG-----FPPAPA--NISTACFSPGVPTAHSNTMPTTQAPLLSREFFYR-- 463
Db 1466 GPGQAAVNAVQRYOVLPRADTLHFATETSTPDGFCSSLSLALSDPEFIQDVELRIM 1525
QY 464 ---EONDKGRSEKFPYSGSSYSTSDSOGLAQHIALTLSPSAHTLDDLHDHPHP 520
Db 1526 PVPQVQDNGNETE---SEQPRESENQEKAEKIDS-----EKDLDDSDDDD 1571
QY 521 EEAELASAMTVHPDMLMDIAHARSRPYRYRSRSRSPPEFR---GQSPTKNVPREEK 577
Db 1572 TEL-LEECITISAMP-----TKSSRKAKKPAOTASKLPLPPVARKKSQLPVYLLLSQNR 1623
QY 578 EREYFRKREVPPPYDIKAYV---GRSVDFRDPFEKERYREWEKRYKGYKGYAVG 634
Db 1624 LQP--QKHVSFTPGDDMPRVYCVGEGTPIINSTATSLSDLT-----IESPNNELAAG 1672
QY 635 AOPRPSANREDSPERLLPLNINSPPTKGRREDIYAAGSHRNKNGATPELSTROSH 694
Db 1673 EGVRGAGQSGEFKRDITP-----TEGRSTDEAOG-----GKTSVYTIPELDN 1716
QY 695 NAKDN-----PKSKEKE-----SENVPG-----DGKGNKKKKHR 723
Db 1717 KAEGDILAECTINSAMPKSKSHKPFYKKTIMDOYQOASASSAPNKNQLDGKKKKKFTSPV 1776

```

```

0Y 724 K-----RNEKEGSESESLFNPLLTETSKRCRSSSIDETKTDTLFLVPSRDA----772
Db 1777 KPLPONTETRYVRKKNADSNNMLNARVVSNDKDSKKQMLKNNSKDFONDKLPNNEDRYRG 18366
0Y 773 -----TPRDEPM-----DAESITFKSVSDKDRKDKPRKVSXDTKRRKSDGA 816
Db 1837 SFADSPHNHTPIEGITPYCFSRNDLSLIDF-----DDDDVDLSRREKALEKRAKEKESEA 18922
0Y 817 TAKDNVLPKSKGPQOEKVDGDRKESRSPSEPLKKAKEATKIDSVK--PSSSSODEK--872
Db 1893 KVTSHTELTJNNOQSANTQAIKAPINRGQPKPILOKOSTFPOSSKODIPDRCAATDEKLO 19522
0Y 873 ---VTGRP-----KRAHSAADTR---ROQOPTRRSKTRVPTS 907
Db 1953 NFALETPPVCFSHNSLSLSDIDQENNNKENEPIKETEPPDSQEPSPKPAQASGYAPKSF 20122
0Y 908 SOKSQPRTRRPRLARKIN-----LI-----ARENNERK 938
Db 2013 HVEITPYCFSRNSLSLSDIDEDLLQDCISSAMKPKKPRRLKGDNEKHSPRMMGIL 20722
0Y 939 -----RKSVDK 945
Db 2073 GEDTLDLKDIQRPDEHGLSPSENFDMWKAIOEGANSIVSLHOAAAAACLSRQASDS 21322
0Y 946 DF-----ESSMKISKEVEGELYKPPPKKMGKDVKELETEREKTI 967
Db 2133 DLSLSLKSGLSPFHLTPDOEKEPFTSNKKGRIILKPEKSTLE--TKLIESESGK-L- 2189
0Y 988 ASSTPAKKIKLNNETREKKIGNAENMSTJREPERKELESTSKIQOEKVAKAKRNVAGSE 1047
Db 2190 -----GCKRYKSLITCKVANSSEIIGOMKOP---LOAMPSTIS---RGRMHIIPGV 2237
0Y 1048 GSSSTLVDTYSTSGSGSPVR-----KSEKTDTKRTVIRKTMEEYNNDMTPAPAEVYIMI- 11022
Db 2238 NSSSS-----TSPVSKKGPIKTPASKSPSBOGATATSPRGAPVSKSELSPARQTSIG 2293
0Y 1103 ---QVPOSKMDKDFESEEEDVTTQPIQSVKPS-----SITKN 1139
Db 2294 GSSKAPRSRSGSRSTSPSPAQOPLSRPIOSPGRNNSISPGRNISPNNKLSQLPRTSSPST 2353
0Y 1140 VTTSPASTAK--YTE-----KESQOPKTIQKLKPEASHELMHQELRSSGASSS 1186
Db 2334 ASTSSSGKMKSTSPGRQMSQONTIKOTGLSKNASSITRSESASAKGLMOMNGAMKK 2413
0Y 1187 -EKRAADREHSGSEKDNPPK-----KSGAQPDKESTYDR--LSBOGHFTLSQSS-- 1236
Db 2414 VELRMSSTSSGSEDSRSERPVLVQSTFIKAPPTLIRKLLESASGESLSPSSRAS 2473
0Y 1237 ETRTSEK-----HESVYSGSSNKKDFTGRDKKVYDSDRYSSSKRDERGE 1281
Db 2474 PTRSOAOTPVLPSPILPMSLSTHSSVQAGWKRFLPNLSPTIEYN--DGRPAKRHD--- 2527
0Y 1282 LAR-RKOSPPRGKESLSG--QKSKLEERLPLPKKGAESKSNSSPPRODKPKHDHAPET 1338
Db 2528 IARSHSSPSPLPINRSGTWKRHRSHKSSSLPVSPTWARTGSSSITLSASESSE----- 2582
0Y 1339 KRPEETKPYDKNGKEREKHAAEARNG 1366
Db 2583 KAKSEDEKHWNISGTYQSKENOVSAKG 2610

RESULT 14
US-08-450-582-7
; Sequence 7, Application US/08450582
; Patent No. 6114124
;
; GENERAL INFORMATION:
;
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSIYU, GEOFF
; APPLICANT: KINZLER, KENNETH
;

```

1 APPLICANT: MARKHAM, ALEXANDER F.
 2 APPLICANT: NAKAMURA, YUSUKE
 3 APPLICANT: THLIVERIS, ANDREW
 4 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 5 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 6 NUMBER OF SEQUENCES: 102
 7 CORRESPONDENCE ADDRESS:
 8 ADDRESSEE: Banner & Witcoff, Ltd.
 9 STREET: 1001 G Street, NW
 10 CITY: Washington
 11 STATE: D.C.
 12 COUNTRY: USA
 13 ZIP: 20001-4598
 14 COMPUTER READABLE FORM:
 15 MEDIUM TYPE: Floppy disk
 16 COMPUTER: IBM PC compatible
 17 OPERATING SYSTEM: PC-DOS/MS-DOS
 18 SOFTWARE: PatentIn Release #1.0, Version #1.25
 19 CURRENT APPLICATION DATA:
 20 APPLICATION NUMBER: US/08/450,582
 21 FILING DATE:
 22 CLASSIFICATION: 435
 23 PRIOR APPLICATION DATA:
 24 APPLICATION NUMBER: US 08/452,655
 25 FILING DATE: 25-MAY-1995
 26 APPLICATION NUMBER: US 08/289,548
 27 FILING DATE: 12-AUG-1994
 28 PRIOR APPLICATION DATA:
 29 APPLICATION NUMBER: US 07/741,940
 30 FILING DATE: 08-AUG-1991
 31 ATTORNEY/AGENT INFORMATION:
 32 NAME: Kagan, Sarah A.
 33 REGISTRATION NUMBER: 32,141
 34 REFERENCE/DOCKET NUMBER: 1107,49964
 35 TELECOMMUNICATION INFORMATION:
 36 TELEPHONE: 202-508-9100
 37 TELEFAX: 202-508-9299
 38 INFORMATION FOR SEQ ID NO: 7:
 39 SEQUENCE CHARACTERISTICS:
 40 LENGTH: 2843 amino acids
 41 TYPE: amino acid
 42 STRANDEDNESS: single
 43 TOPOLOGY: linear
 44 MOLECULE TYPE: protein
 45 HYPOTHETICAL: YES
 46 ANTI-SENSE: NO
 47 US-08-450-582-7

[illegible]

| | | | | | | | |
|----|------|--|---|--|-------------|----------------------------|-----|
| 0y | 421 | PRG----- | ----- | FPAPA-- | ---NISTC | SPBGPVTAHSMTPQOAPLLSREPR-- | 463 |
| Db | 1466 | GPKOAVNAAYQROYVLP | DDTLHLHATATSTPDGSCSSLSALSDEPFIQKVLEIRIM | 1525 | | | |
| 0y | 464 | ---EOND | KGRESFPYSGSSYSSTDSOGLAHIALTLSPSAHTLDDLHDHBP | 520 | | | |
| Db | 1526 | PRVGENONGNETE----- | ---SQPKRESHENGEKAEKTTDS----- | EKLDDSDDD | 1571 | | |
| 0y | 521 | EEAFARSAMTIVHMPDLMDLAHAHSRPPYRRYRRSRSPER-- | ---GSPFKRVVPEEK | 577 | | | |
| Db | 1572 | IEI-LEECIIISAMP----- | ---TKSSHRAKKPAQOTSKLPPYARKPSOLPYUKLLPQNR | 1623 | | | |
| 0y | 578 | EREFRNFREVPPEYDIKAY-- | ---GSRVFRDPFEKERYREMERKYREMYKXYKGYAVG | 634 | | | |
| Db | 1624 | LQP---QHV | SFTPDGDDPRVYCEYGPPIFSTATSLDIT----- | IESPNEILAAG | 1672 | | |
| 0y | 635 | AQPPASNREDFSPERLLPLTINRSPPTGRREDYLAAGSHRNRLGNYPEKLISTDSH | 694 | | | | |
| Db | 1673 | EGVAGGAQSOSEFKEKROTI P----- | ---TGRGRIDEAQG----- | GKTSSVTIPELDNN | 1716 | | |
| 0y | 695 | NAKN----- | PKSEKE----- | SENVPG--- | DKGKHKHKHR | 723 | |
| Db | 1717 | KAEBGDIILAECSINAMKGGSHKPFVRKKIMDOYQOASASSAPKKNOLDKRRKPTSPV | 1776 | | | | |
| 0y | 724 | K----- | RRNEKGESESRFLNPELLETSRKCGSSGIDETKTDTLFLVPSRDA--- | 772 | | | |
| Db | 1777 | KPIQNTETRYRRRKNAADSNNINIAERVSNDKDKOKNLAKNSDOKFMDKLPNNEDRYRG | 1836 | | | | |
| 0y | 773 | ----- | TPVDEPM----- | DAESITPFKSVSDKDKREKDKPRKVSDDTKRRKSDGSA | 816 | | |
| Db | 1837 | SFAPDSPHHTPIEGPYPCFSRNDLSISLDF----- | DDDDVDLSREKALERAKENKESEA | 1892 | | | |
| 0y | 817 | TAKDNYLAKSPKGOEYVDDREKSPRSEPEPLKKAEEKITDYSK-- | PSSSOKDEK-- | 872 | | | |
| Db | 1893 | KVTSHTELTSNOOSANKTQOIAQIPNRQOPKPILOKOSTPPOSKDIPDGAATDEKLO | 1952 | | | | |
| 0y | 873 | ---VTGP----- | ---RAHSASAKDTR--- | ROSQPTRRSKRTVPTS | 907 | | |
| Db | 1953 | NFAIENTPVCNSHNSLSLSLSDIDQENNNKENPEIKETPPDSQCEPKPQASGTAIRSF | 2012 | | | | |
| 0y | 908 | SQKSPQVATRRPRSLRKINY----- | LI----- | AREKNEREK----- | 938 | | |
| Db | 2013 | HVEDPVCFSNNSLSLSLSIDSEDLLOECISAMPKKKPKRLKGDNKHSPRMGGIL | 2072 | | | | |
| 0y | 939 | ----- | ----- | IKKSYDK | 945 | | |
| Db | 2073 | GEDLTLDKDIQRPDSEHGLSPSENEFMWKAIOEGANSIVSSLHOAAAACLSROASSDS | 2132 | | | | |
| 0y | 946 | DE----- | ---ESSMKISVCEGEIYKPSPKRKMGGVDEKLEPRPEKRI | 987 | | | |
| Db | 2133 | DSILSLKSGISLSPFHULTPDQOEKPTTSKNGRILLKPEEKSTLE- | IKKISESEKGIK- | 2189 | | | |
| 0y | 988 | ASSTTPAKIKLNETEKKIKIGNAENASTTPEPEKLESTSSKIKOEKYGAKRRKRVAGSE | 1047 | | | | |
| Db | 2190 | ----- | GKKVYKSLITLTKVANSNSLSQMMQP--- | LQAMMPsis--- | RGRMIHITPVR | 2237 | |
| 0y | 1048 | GSSSTLVDTYSTSGSGSPVR--- | KSEKTDTKRTVIKTMEBYNNDNTAPAEVYIIM- | 1102 | | | |
| Db | 2238 | NSSSS----- | TPSVKSKPPPLKTPASKSPSEGOATATSPRGACPPYSKELSIPVARTSOIG | 2293 | | | |
| 0y | 1103 | ---QVP | SKMKXKDPSESEEDVKTQPIQSVGRPS----- | SIINK | 1139 | | |
| Db | 2294 | GSSKAPSRHSGRSDTSPRPAQOPLSRPIQSPGNSSISPRNGNISPPNKLISQLPRTSPST | 2353 | | | | |
| 0y | 1140 | VTTKRSATAK---YE----- | ---KESQEPKLOKLPREASHELMOHELSSKGSASS | 1186 | | | |
| Db | 2354 | ASTKSSGSGKASITSYSPGRHMSQOOLTKQTGLSKNASSITPRSESASAKGLNOMNGGAKMK | 2413 | | | | |
| 0y | 1187 | -EKGRAKREHSGSEKDNPK--- | RKSGAQDPKESTVDR--- | LSQEGHETLQSOSK--- | 1236 | | |
| Db | 2414 | VELSMTSSTKSSGSESDRSEKRPVLYVROSTTIKRAPETILRRKLEASAEPSLSPESSRRAS | 2473 | | | | |
| 0y | 1237 | ETRTSEK----- | ---HESVAGSSKNDFTPGRKDKVYDSDRYSSSKRRDERGE | 1281 | | | |

```

Db 2474 PTFRQAOATPVLSLPLPMSLSTHSSVQAGGWRKRLPNNLSPTIEYN--DGRPAKRD---- 25237
      ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :
Oy 1282 LAR-KDUSPPRGKESLUG--QKSKLAEERDLPKKAESKSSSSPPRDKKRPNDKAPYET 1338
      ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :
Db 2528 IARSHSESPSLPLNRSGTWRKREHSHSSSLPVSPTWRTGSSSILASSESE----- 2582
      ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :
Oy 1339 KRCEETKPYVDKNSGKEREKHAALRNAG 1366
      ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :
Db 2583 KAKSEDEKHWNSIGTKQSEKENVSAKG 2610
      ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :  ||:  :

```

```

15
US-08-821-355A-7
: Sequence 7, Application US/08821355A
: Patent No. 5851775
: GENERAL INFORMATION:
: APPLICANT: Barker, Nick
: APPLICANT: Clevers, Hans
: APPLICANT: Korinek, Vladimir
: APPLICANT: Morin, Patrice
: APPLICANT: Kinzler, Kenneth
: APPLICANT: Vogelstein, Bert
: APPLICANT: Sparks, Andrew
: TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
: TITLE OF INVENTION: Interact to Prevent Cancer
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Banner & Witcoff, Ltd.
: STREET: 1001 G Street, N.W.
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20001
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: Fastseq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/821,355A
: FILING DATE: 20-MAR-1997
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Kagan, Sarah A 32,145
: REGISTRATION NUMBER: 1107.05064
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-508-9100
: TELEFAX: 202-508-9299
: TELEX: 97430 BMB UT
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2973 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: NO. 5851775e
US-08-821-355A-7

```

Query Match 3.78; Score 274; DB 2; Length 2973;

Best Local Similarity 18.5%; Pred. No. 1.1e-09;
Matches 282; Conservative 188; Mismatches 612; Indels 446; Gaps 61;

[illegible]

```

Db 1247 ---ATCKSSINQETIQYCVEDTPICFSCSSLSLSSADELIGNOTQEOADANLTLQ 1303
Qy 275 ITALMEEG-----VPGTSP-----WNSTFVQSLHGLIPIITG---PVR 312
Db 1304 IAEIKETIGTRSAEDPVEVAVSQHPRTKSSRLQSSLSSESAHKKAVERSQAKSSDK 1363
Qy 313 INARPGGGRGWESHNKLGIVSPPOQIRGERSCYRSINNGRHHSRORTQSPSLPA 372
Db 1364 SGAQTP-----KSPREHYVOETPLMFESRCTSVSSLDSEFESRIASSVQS 1407
Qy 373 TPCEVPV-----PPPLYPHPHTLPLPGVPPQFSP---QFSPSPPTAGYVSP 420
Db 1408 EPCSGMVGIISPSDLPSDLPQGTMPSPSKTTP---PPQGTQTKREYVKNNAPTAKRES 1465
Qy 421 PPG-----FPRAA---NISTACSPGVPTAHSTMTPTQAPLISREFFYR--- 463
Db 1466 GPKQAAVNAVQOVLEDAOTLLHFAFESTPDPGFCSSLSLALSLDPEFIOKQVLEKIM 1525
Qy 464 ---BONDGRSKRPYSGSVSRSSYTSSQGLAQHIALTLSPSAHTLDDLHDPHP 520
Db 1526 PPOVENDGNETE-----SEQPKESNEQKEAKETIDS-----EKDLDDSDDD 1571
Qy 521 EEAERASAMIVHMDLMDIAHARSPPYRRYRSRSRSPPEFR---GOSPTKRANVPREK 577
Db 1572 IEL-LEECIIISAMP-----TKSSRKAKKPAQTASKLPPVARKPQSLPYKLLPSQNR 1623
Qy 578 EREFNRKREVPYPPYDIAYV---GRSVDFNDPFEKERYREMEKRYREMEKYGYAVG 634
Db 1624 LQP---OKHVSFTPGDDMPRYCVECTPIINFSTANSLDLT-----IEPPNELAAG 1672
Qy 635 AOPPSANREDFSPERLLPLINRSPFTRGRREDYAAQSHRNRLGNGNYPEKILSTRDSH 694
Db 1673 EGVNGGAOSGEFEKROTIIP-----TEGRSTDEAQ-----GKTSSVTIPELDN 1716
Qy 695 NAKDN-----PKSKEKE-----SENVPG---DGKGNKKKKHR 723
Db 1717 KAEGBDILAEICINSAMPKRGSHKPPRVKKINDQVOQASASSAPKQNLDDGKKKPTSPV 1776
Qy 724 K-----RNEKEGESEFNLPELLETSKRCGSSGIDETKTDTLFLVLRSDA--- 772
Db 1777 KPDPONTETRTVRKKNADSKNNLNAERFVSNDKSKKONLKNNSKDFPNNEDRYVG 1836
Qy 773 -----PPVRDEPM-----DAESITPKSVSDKDKREKDKPKVKSDDTKRKSDGSA 816
Db 1837 SFADSPHHYPIEGTPICFSGRNDLSLDR---DDDDVDLSREKAELRAKENKESEA 1892
Qy 817 TAKDNVLPKSGPQEKYDGDREKSPRESEPLKKAKEATKIDSVK---PSSSSQKDEK-- 872
Db 1893 KVTSHTELTSNQOSANKQATAKQPINRGOPKPILOKOSTPPOSSKIDIPRGAATDEKLQ 1952
Qy 873 ---VTGTP-----RKASHKSAKQTR---ROSQPTRRSKRTVPKTS 907
Db 1953 NFAIENTPVCFSHNSLSLSDIDQENNNKENEPIKETEPDPOGEPSKPOASGYAPKSF 2012
Qy 908 SOKSOPVTRRPRSLRKINY-----LI-----AREKNEREK----- 938
Db 2013 HVEDTPVCFKSNSSLSLSDIEDLLQECISSAMPKKKKPSRLKGDNKHSPPRMGGIL 2072
Qy 939 -----RKSSYDK 945
Db 2073 GEDLTLDKDIORPDEHGLSPDSENFDMKAIQEGANSIVSSLHQAAAAACLSRQASSDS 2132
Qy 946 DF-----ESSMKIKVEGTEIYKSPKRMKGDYKLETPPEKDKI 987
Db 2133 DSIILSKGSIIGSPFHLJTPQOEKPFITSNKGPRILKGEKSTLE--TKKIESESGIK- 2189
Qy 988 ASSTPAKKIKILNRETGKIGNAENASTKEPSEKLESTSSKIKQEKVKGAKKRVAGSE 1047
Db 2190 -----GKKVKYKSLITGKVRNSSEISGOMKOP---LOANMPSIS---RGRTHIHPGVR 2237
Qy 1048 GSSSTLVDTYSTSTSGSPVR---KSEKTDTKRTVIKTMEYVNNNDNTAPAEDVIIMI- 1102

```

```

Db 2238 NSSSS-----TSPVSKKGPPLKTPASKSPSEQATATTPRGAKPSYKSELSPYARQTSQIG 2293
Qy 1103 ---QVPQSKMKDKDFEESPEEDVKTTOPIQSVGKPS-----SIITKN 1139
Db 2294 GSSKAPSRSGSHDSTPSKRPAAQOPLSRPIQSPERNISIPGRNGISPPNKLSQLPRTSPST 2353
Qy 1140 VTTKPSATAK--YTE-----KESDQPEKLOKLPKEASHLMOHELSSKGSASS 1186
Db 2354 ASTKSSGSGKMSYTPGQMSQONLTQOTGLSKNASSIIPRSRSASKGJLQNMNNGAKKK 2413
Qy 1187 -EKGRADREHSGSEKDNPKD---RKSGADPDKESTVDR--LSQGHFKTILSQSK--- 1236
Db 2414 VELSPMSSTKSSGSESDSERPVILRQSTFIKAPSPILRRKLEBSASFESLSPSRPAS 2473
Qy 1237 ERTSEK-----HESVRSSNKNDFTPGDKDKYDSDRYSKRRDERGE 1281
Db 2474 PTRSQAQTPVLSPLPDMSLSTHSSVQAGGWKRLPNNLSPTLEYN--DGRPAKRHD--- 2527
Qy 1282 IAR-RKDSPPRGKESLSG--OKSKLREERDLPKKAESKKSNSSPPRDKKPHDKAPYET 1338
Db 2528 IARSHSESPSLPINRSGTWKREHSHSSSLPRVSTWRTGTGSSSILNASSPSE----- 2582
Qy 1339 KRPECETRPVDKNSGKEREKHAAREANG 1366
Db 2583 KAKSEDEKHVNSISGTQSKENQVSAKG 2610

```

Search completed: September 12, 2002, 19:10:23
Job time: 11187 sec

.....

Ratio: 4.976 Gaps: 9
Percent Similarity: 92.703 Percent Identity: 90.966

alignment_block:
US-09-811-045a-1 x MMU28789 ..

Align seg 1/1 to: MMU28789 from: 1 to: 5191

5 LysAspProAsnMetLysGlyAlaMetLeuThrAsnThrGlyLysTyrAl 21
|||||
2 AAAGATCCCTAAATATGAAAGGTGCAATGCTTACAAACACATGGAATAATATC 51
21 alleProThrIleAspAlaGluAlaTyrAlaIleGlyLysLysGluLysP 38
|||||
52 AATACCAACTATATGATGCAGAGGCCCTATGCATCGGGAAGAAAAGAAAC 101
38 robProPheLeuProGluGluProSerSerSerSerGluLysAspPro 54
|||||
102 CACCCCTTCTACAGAGAGCATCATCTTCAAGAGAAGATGATCCT 151
55 IleProAlaGluLeuLeuLysLeuIleCysLysAspIleMetThrAspAl 71
|||||
152 TTCGCCAGACGAGCTCTGTGCTCATCTGCAAGACATCATGATGATGATGC 201
71 aValValIleProCysCysGlyAsnSerSerCysAspGluCysIleArgT 88
|||||
202 TGTGGTCATTCCTGCTGCTGGAAACAGTTACTGTGATGATGATATAGAA 251
88 hThrLeuLeuGluSerAspLysHisThrCysProThrCysHisGlnAsn 104
|||||
252 CAGCACTCTTGAGTGCAGATGAAACATACATATGTCACACATGTCACCAAAAT 301
105 AspValSerProAspAlaLeuIleAlaAsnLysPheLeuArgGlnAla 121
|||||
302 GAGTTTCTCCGATGCTTAAATGCAACAAAGTTTACGACAGCTGT 351
121 LysAsnPheLysAsnGluThrGlyTyrThrLysArgLeuArgLysGlu 138
|||||
352 TAATTAATCTTAAATGAAACTGGCTATATCAAAACGACTACGAAACAGT 401
138 euProProPheLeuPheLeuValProProArgProLeuSerGlnArg 154
|||||
402 TACGGCACCCACCCACCCAGTACACCAACAGACACACTCATGCAGCGG 451
155 AsnLeuGlnProArgSerArgSerProIleLeuArgGlnGlnAspPro 171
|||||
452 AACCTCAGCCTCTGATGAGATCTCCCATATCATCAAGACAGAGATCCTCT 501
171 ValPheArgTyrThrValSerProThrCysSerAspThrLysThrAlaG 188
|||||
502 GATGATTCAGTGACATCGTCTCCGCTCACTCA..... 535
188 LysCysSerAspSerGlyThrLeuSerArgLeuProAlaProSerIle 204
|||||
536GCTCCCTCTATA 547
205 SerSerLeuThrSerAsnGlnSerSerLeuAlaProProValSerGln 221
|||||
548 TCTTCAATTAATCAATTCATCTGCCGCGCTCTTCTGTCTGTGAAA 597
221 nProSerSerAlaProAlaProValProAspIleThrAlaThrValSer 238
|||||
598 TCCATCTTGGCCCGACGCTCCAGTACCTGATATACCTGCAACCGTGTCTA 647
238 LeSerValHisSerGluLysSerAspGlyProPheArgAspSerAspAsn 254
|||||
648 TATCATCCACCTCAGAAAATCGATGAGACTTTTGGGATTTCTGATAAT 697
255 LysLeuLeuProAlaAlaAlaLeuThrSerGluHisSerLysGlyAla 271
|||||
698 AATTAATTGCGACCTGCGCCCTTACATCATGAAACATTTCAAAGGAGCCTC 747
271 rSerIleAlaIleThrAlaLeuMetGluGluLysGlyValProGlyThrS 288

|||||
748 TTCAAATGCTATTACGCTCTTATGSAAGAAAAGGGCTA.CCAGGTACCA 796
288 erProTrpAsnSerIlePheVal.GlyGlnSerLeuHisGlyGlnLe 304
|||||
797 GTCCCTTGGAACCTCCATCTTGTGGGACAGTCATATATACATGAGACAGTT 846
304 utLeProThrThrGlyProValArgIleAsnAlaAlaArgProGlyGlyG 321
|||||
847 GATTCACCAACTGGCCAGTAAGATCAATGCTGCTGCCAGGTGGTG 896
321 LysArgProGlyTyrPheLysSerAsnLysLeuGlyTyrLeuValSerPro 337
|||||
897 GCGGGCAGAGCTGGGAGCATTCACAACACTGGGTAAGTTCCTCA 946
338 ProGlnGlnIleArgArgGlyGluArgSerCysTyrArgSerIleAsnAr 354
|||||
947 CCACAGCAAAATTAGAAAGAGAGAAAGACCTGTTACAGAAATATAAACCG 996
354 gGlyArgHisHisSerGluArgSerGlnArgThrGlnSerProSerLeuP 371
|||||
997 CCGGGCAGACACACAGGGAAGATCACAGAGACTCAAGGCCCATCACTTC 1046
371 roAlaThrProCysPheValProValProProProLeuTyrProPro 387
|||||
1047 CAGCACTCCAGCTCTTGTGCGCGTTCCACACACTCTTGTATGCGGCT 1096
388 ProProHisThrLeuProLeuProProGlyValProProGlnPheSe 404
|||||
1097 CTCCCAATACACTTCTCTCTCCAGGTGTAACCTCCCCACAGTTTTC 1146
404 rProGlnPheProSerSerGlnProProThrAlaGlyTyrSerValProp 421
|||||
1147 TCCTCACTTCTCTCTGCGCAGCTCCACAGCAGGATATAGTCCCTC 1196
421 roProGlyPheProProAlaProAlaAsnIleSerThrLacysPheSer 437
|||||
1197 CTCACAGGTTTCCACAGCTCCGCTCCATATATCAACACTTGGATATCA 1246
438 ProGlyValProThrAlaHisSerAsnThrMetProThrThrGlnAlaPr 454
|||||
1247 TCAGAGATGCAGACTCCCATTCAAATACATCCCTACAAACAAACACCC 1296
454 oleuLeuSerArgGluGluPheTyrArgGluGln..... 465
1297 TCTTTGTCAGGGAAGAAATTCTATAGAGACAAACAGCGCTAAAGGAAG 1346
466AsnAspLysGlyArg 470
1347 AGCAAAAAGAAAAGTCCAAAGCTGATGAGTTTACAATGATTTTGGCTAAG 1396
471 Glu..... 471
1397 GAATGTATGGAATACAAAAGATTCAAAAAGAGCGTAGGCGCTCATTTTC 1446
472SerLysPheProTyrSerGlySerTyrArgSerSerTyrT 487
1447 CAGGTATAATCTCCCTATAGTGGTATCGTATTAAGAAAGTTCATACA 1496
487 hrAspSer.SerGlnLysLeuAlaGlnHisIleHisAlaLeuThrLeuSe 503
|||||
1497 CTATTTCMAAGTCAAGTCTGGCTCAACAGCTTACAGCTTATCTGTGCG 1546
503 rProSerAlaAlaHisThrLeuAspLeuHisAspHisProHisProp 520
|||||
1547 TCTTCAAGCGCTCACATTTCTCGCTCTATTCACGATCACCCCAATACCC 1596
520 roGluGluAlaGluAlaArgSerAlaMetIleValHisMetProAspLeu 536
|||||
1597 CAGGAAGGAGAGGAGGAGGAGCCGCAATTACGCTTACAGGTCACAGATTC 1646
537 MetAspIleAlaHisAlaArgSerArgSerProProTyrArgTyrArg 553
|||||

1647 ACGGATACAC .CGATCTAGGTCAAGGTCACTCCCTATAGACGATACA 1695
553 gSerArgSerArgSerProGluPheArgGlyInSerProThrIysA 570
1696 CTCACGGTCCAGATCTCTCAAGCATTTAGGGGACAGTCTCCCTAAC 1745
570 rGAsnValProArgGluGluGluArgGluTyrPheAsnArgTyrArg 586
1746 GCATGTACCTCAAGAGAAACAGAGCGGTGATTTTAATAGATACAGA 1795
587 GluValProProProTyrAspIleValaTyrTyrGlyArgSerValas 603
1796 GAAGTTCACCCCTTATGACATCAAGCCATTATGCGGAGTGTGCA 1845
603 pPheArgAspProPheGluLysGluArgTyrArgGluTyrPgluArgIysT 620
1846 CTTTAGAGACCATTTGAGAAAGAACGCTACGGGAATGGGAAAGAAAT 1895
620 yTArgGluTyrPyrGluLysTyrTyrIysGlyTyrAlaValaGlyAlaGln 636
1896 ACCGAGGTGTATGAGAAAGTACTACAAAGGGTACGCGGTGGAGCTCAA 1945
637 ProArgProSerAlaAsnArgGluAspPheSerProGluArgLeuLeuPr 653
1946 CTTAGACCTCAGCCAAATAGAGAGACTTCTCCAGAGAGACTCTTACC 1995
653 OLeuAsnIleArgAsnSerProPheThrArgGlyArgArgGluAspTyrA 670
1996 TCTTAATATCGAAATTCACCCCTTACAGAGGCGGAGAGACGATAG 2045
670 IaAlaGlyInSerHisArgAsnArgAsnLeuGlyIysAsnTyrProGlu 686
2046 CTGCTGGCAAGTCATGAAATGAAATCTAGGTGGCACTATCCAGAA 2095
687 LysLeuSerThrArgAspSerHisAsnAlaLysAspAsnProLysSerLys 703
2096 AAGCTTTCACAGAGGACGACATGACAAATGCAAAAGTAATCCAAATGCAA 2145
703 sGluLysGluSerGluAsnValProGlyAspGlyLysGlyAsnLysHisL 720
2146 GGAGAGGAGAGTGAATGTTCCAGAGAGCGCAAGAGGAAACAGCAT 2195
720 ySLysHisArgLysArgArgAsnGluGluLysGlyGluGluSerGluSer 736
2196 AGAAACAGAGAAACGAGAAAGGGGAGAGAGAGTACAGAGC 2236
737 PheLeuAsnProGluLeuLeuGluTyrSerArgLysCysArgGlySerSe 753
2237 TTCCTGAACCCAGAGCTACTGAGACGCTCTAGAAATGCAGGAAATGCTC 2286
753 rGlyIleAspGluThrLysThrAspThrLeuPheValLeuProSerArgA 770
2287 AGGATTGATGAACGAAAGACAGATACACTGTTGTTCTCTCCAGACAGAG 2336
770 sPAspAlaThrProValaArgAspGluProMetAspAlaGluSerIleThr 786
2337 ACGATGCTACACCTGTTAGGATGAGCCAATGACGCAGAATCGATCCT 2386
787 PheLysSerValSerAspLysAspLysArgGluLysAspLysProLysVa 803
2387 TTCAAAGTCAGATCTGACAAAGACAAAGAGGAAAGGATTAACCCCAAAAGT 2436
803 LysSerAspLysThrLysArgLysSerAspGlySerAlaThrAlaLysL 820
2437 AAAAGTGCACAAAGACAAAGGAAAGTGAGGGTCTCTACAGCCAAAGA 2486
820 ySAspAsnValLeuLysProSerLysGlyProGlnGluLysValaAspGly 836
2487 AAGACAAATGTTTAAACCTCTTAAGAGACCTCAAGAAAGATGATGGA 2536
837 AspArgGluLysSerProArgSerGluProProLeuLysLysAlaLysGlu 853
2537 GACCGTGAAGTCTCTCGGTGTGAGCGCCACTCAAAAAAGCCAAAGA 2586

853 uGluAlaThrLysIleAspSerValLysProSerSerSerSerGlnLysA 870
2587 GGAGGCTACAAAGATGACTCTGTAAACCTTCCTCTTCTCAGAAAG 2636
870 sPgluLysValaThrGlyThrProArgGlySalahisSerLysSerAlaLys 886
2637 ATGGAAGGTCTACGTGAACCCCTGAAAGCCCATTTCTTAATCTGCAAAA 2686
887 Asp . ThrArgArgGlnSerGlnProArgThrArgArgSerLysArgThrV 903
2687 GAACACCAAGAGGCAAAAGCCAGCCAAAGAGAGAGAGTCAAAAAAGACTG 2736
903 aLProLysThrSerSerGlnLysSerGlnProValaArgThrArgArgPro 919
2737 TTCCAAGACATCAAGTCAGAAAGCCAGCCAGTACAGAGAGAGGCCCA 2786
920 ArgSerLeuArgLysIleAsnTyrLeuIleAlaArgGluLysAsnGluArg 936
2787 AGAAGCCTGAGAAATTAATCTACTGTATGCAAGGAGAGAAACGAAAG 2836
936 gGluLysArgLys . LysSerValAspLysAspPheGluSerSerSerMe 952
2837 AGAAAGCGGAGAGAAAGAGTGTAGATTAAGATTTGTGCTCTTCGAAT 2886
952 tLysIleSerLysValaGluGlyThrGluIleValaLysProSerProLysA 969
2887 GAAATCTCTTAAGTAGAAGGACAGAAATAGTAAACCATCACCAAAAC 2936
969 rGlyLysMetGluGlyAspValaGluLysLeuGluArgThrProGluLysAsp 985
2937 GGAATAGGAAAGGTGATGTTGAAAGACTCGAAAGACCCCGAAAGAGAC 2986
986 LysIleAlaSerSerThrThrProAlaLysLysIleLysLeuAsnArgGlu 1002
2987 AAGATTGCATCATCACTACCTCCAGCCAAATAATCAAACTCAACAGAGA 3036
1002 uThrGlyLysLysIleGlyAsnAlaGluAsnAlaSerThrThrLysGluP 1019
3037 AACCTGAAAAAAATTTGMAATGCAAGAAATGCAATCTACATCAAAAGAAC 3086
1019 rOserGluLysLeuGluSerThrSerSerLysIleLysGlnGluLysVal 1035
3087 CCTCTGAAAAATTTGAGTCAACATCTACCAAAATCAAAACAGAAAAAGTC 3136
1036 LysGlyLysAlaLysArgLysValaGlySerGluGlySerSerSerTh 1052
3137 AAGGAAAGGCCCAAGCGAAAGTAGCTGGTGGAGGAGCTCCACCTCCAC 3186
1052 rLeuValAspTyrThrSerThrSerSerThrGlyGlySerProValaArgL 1069
3187 GCTGTGATTTACACAGTACAAAGTTCAACCTGAGGAGAGTCTGTGAGA 3236
1069 ySerSerGluGluLysThrAspThrLysArgThrValIleLysThMetGlu 1085
3237 AATCTGAAGAAAGACACAGATCAAAAGCAAGACAGTCATTAAACATATGAG 3286
1086 GluTyrAsnAsnAspAsnThrAlaProAlaGluAspValIleIleMetI 1102
3287 GAATATATATATGATTAACACAGCTCTCTGACAGATGTTAATAATATGAT 3336
1102 eGluValaProGlnSerLysTrpAspLysAspAspPheGluSerGluGlu 1119
3337 CCAGGTCTCTCAGTCCAAATGGAATAAGATGATTTGAGTCAAGAGAG 3386
1119 LuAspValLysThrThrGlnProIleGlnSerValGlyLysProSerSer 1135
3387 AAGATGTTAAACCAACACACATCTACAGAGTGTAGGAAACCAATCGAGT 3436
1136 IleIleLysAsnValThrThrLysProSerAlaThrAlaLysTyrThrGlu 1152
3437 ATTATAAAAATGTCACACTACTAAGCCATCGGCTACGGCTAAGTACACCGA 3486

1152 uylsglulsergluInProglulysleuInlylsleuProlysgluAlas 1169
 1169 |||||||
 3487 GAAGGAAACGAGCAGCCGAGAACTGCAGAACTCCCAAGAGAGCCA 3536
 1169 erHisGluLeuMetGlnHisGluLeuArgSerSerlysglySerAlaSer 1185
 1185 |||||||
 3537 GCCACAGCTGATGACGACGAGCTCAGCAAGCCCAAGAGGAGTGGCTCC 3586
 1186 SerGluLysGlyArgAlaLysAspArgGluHisSerGlySerGluLys 1202
 1202 |||||||
 3587 AGTGAAGAGGAGAGCAGCAGGAGCCGAGCACTCGAGGTCCGAGAGGA 3636
 1202 pAsnProAspLysArgLysSerGlyValaGlnProAspLysGluSerThyV 1219
 1219 |||||||
 3637 CAACCTGACAGAGAGAGAGCGGTCGCCAGCAGCAAGAGAGACACTG 3686
 1219 aLysArgLeuSerGluGlnGlyHisPheLysThrLeuSerGlnSer 1235
 1235 |||||||
 3687 TGGACCGCTGAGTGAGCAGGAGCATTAAAGCTCTCTCTGCTCTCC 3736
 1236 LysGluThrArgThrSerGluLysHisGluSerValArgGlySerSer 1252
 1252 |||||||
 3737 AAAGAGACAGAGAGCTTCAGAGAGCAGAGTCTGTCGTGCTCTCAAA 3786
 1252 nLysAspPheThrProGlyArgAspLysLysValAspLysSerArg 1269
 1269 |||||||
 3787 TAAAGCTTCACTCTGCTAGAGACAGAAAGTGGACTACGAGACAGAGG 3836
 1269 sPlySerSerSerLysArgArgAspGluArgGlyGluLeuAlaArgArg 1285
 1285 |||||||
 3837 ATTATTCAGATTCCAAAGAGAGAGAGAGAGTGAATTAAGAGAGAGA 3886
 1286 LysAspSerProProArgGlyLysGluSerLeuSerGlyGluLysSer 1302
 1302 |||||||
 3887 AAAGACTCTCCCTCCCGGAGCAAAAGTCTCTGTCGGCAGAGAAAGCAA 3936
 1302 sLeuArgGluGluArgAspLeuProLysGlyAlaGluSerLysLys 1319
 1319 |||||||
 3937 GCTGAGGAGAGAGAGATTACCTAAAAAGGGGCGGAGTCAAAAAAAA 3986
 1319 erAsnSerSerProProArgAspLysLysProHisAspHisLysAlaPro 1335
 1335 |||||||
 3987 GAAATCTGAGCCCCCAAGAGACAAAAAGCCATGATCAATAAGCCCC 4036
 1336 TyrGluThrLysArgProCysGluGluThrLysProValAspLysAsn 1352
 1352 |||||||
 4037 TACGAAACTAAAGCCCATGTGAAGAGACAAAGCCGTGATTAATACTC 4086
 1352 rGlyLysGluArgGluLysHisAlaAlaGluAlaArgAsnGlyLysGlu 1369
 1369 |||||||
 4087 TGGGAAGAGAGCGGAGAGACATGCTCTGAAAGCTCGCAATGGAGAGAGT 4136
 1369 erSerGlyAlaAsnCysHisValTyrLeuThrArgGlnThrLeuProTr 1385
 1385 |||||||
 4137 CCAGTGTGGCAAACTGCATGATACCTAACCCGAGACCTCCAG 4186
 1385 parGArgSerTrpLeuLeuGlyArgTrpArgAlaProSerSerArg 1402
 1402 |||||||
 4187 GAGAAAGAGCTGCTGCTGGCAGGTGAGAGAGCGCGTCAAGCCGAA 4236
 1402 snProSer 1404
 1402 |||||||
 4237 ACCCCAGC 4244

seq_name: gb_pr:AF352051
 seq_documentation_block:
 LOCUS AF352051 5376 bp mRNA linear PRI 20-SEP-2001
 DEFINITION Homo sapiens proliferation potential-related protein mRNA, complete cds.
 ACCESSION AF352051
 VERSION AF352051.1 GI:15705402
 KEYWORDS

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 5376)
 AUTHORS Camargo,A.A., Moreira,E.S. and Simpson,A.J.G.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2001) Laboratory of Cancer Genetics, Ludwig
 Institute for Cancer Research, Rua Antonio Prudente 109 4th floor,
 Sao Paulo 01509-010, Brazil
 FEATURES
 source
 1..5376
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 92..4942
 /note="Similar to mouse proliferation potential-related protein"
 /codon_start=1
 /product="proliferation potential-related protein"
 /protein_id="AA105625.1"
 /db_xref="GI:15705403"
 /translation="MGIKTLNLVLGLRALPEPEVMEVYKDPNMKAMLTNGKVAL
 PTIDAEVYALGKEKPEPLPEPESSSEDDPIDELCLICXIDIMTDAVYIPCCGNS
 YDCETRALLESDEHPCPTCHONVSPDALIANKFLROAVNNKNETGYTRLRKGL
 PPPPIIPPPRLIQRNLIQPLMRSPISROODPLIPTYSSSTHPAPISISITSMOSL
 APYSGNPSRPAPVPITATVTSISVSEKSDGPFRRSDNKTILPAALASHSGTSS
 IATLMEKGYQVPLGTPSLGQSLLHGLIPTGVRINTARPGGGRGWHSNK
 LGYLVSPPOQIRBERSCYRSINGRHHSESOPTQPSLPATVFPVPPVPP
 PPHLPPLPPGPPQFPQFPQPGPPAGYVPPGPPAPAMNSTPWSVGVYAHNS
 TTPTOAPPLSREEYREORRLKEEKKSLEDEFTNDEFAKLEMYKIOKERSS
 NYSRSNGSGYSRSSYTKSRSGSTSRYSRSPSSHSYRSRSPYPRGGRKR
 NRSRSNGHYSRHSRSPYRRYRHSRSRQARQGSPPKRNYPQGETEYENRRR
 EVPPYDMKATYGSVDPDRPEKERYREWRKATREWEYKYVAAAGAOFRSANE
 NFSERPLPLMIRNSPFTGRREDYVGQSHRSNLSINPEKLSADGHAKNTLS
 KEKESENAPGDKGNKHKRRKRGSESEGLNPELLETSRKSREPTGVENKTDL
 FVLPSRDATVPDRPDMAESIIFKSVSEKRDMDPKAKGDTKRRKNDGSAKKE
 NIVKPAKGQKGVGERSPSRSPPIKKAKEELPKDNTSSSSSGNIDKGTTPRK
 AHSKAKEHOETPKYKEKYKDYKSKYKSKLTTPKEKAKKPKMKKPLDNKGEKK
 RKTLEKGYKDFESSKISKLEVELEYKSPKRMEDTPEKMDTPBEKDISLARA
 KRIKLNRETKRIGSTENISNTRKEPSEKLESTSVKVOEKVGRKVTGTESSST
 LVDTYSTSTGSPVRKSEKTDTRKVIKIMBYNNDNPAADVILMIOVPSKMD
 KDDESEEDVSTOPISSVGKPAVAKINVTKSNINVKYPEKSESEKIOKPTKV
 SHELIOHEVKSRSASSEKGTCTDRODVYLEKNEPKRKSSTOPEKESNDRLENG
 NFKLSOSKKAERTSDKHDSRTASNNKDFPNRDKTDYDREYSSSKRRDEKELTR
 RKDPSRKKBDSAGCKNPPREEDLPKGTGDSKSSNSBDRKPHDKATYTTRK
 NEETKSVKNPCKRKHVLEARNKSSGKRLLYILNPETOYEKQITGQIDKSTV
 KPKPOLSHSSRLSDLTRETDPAFEDPDYNSDESNVSYEESSGNISGDLDKIV
 EKAKESLDTAIVOVGISRNOSHSPPSVSPSRSHSPSGOTRSHSSASMEOSDSK
 KKKKEKKKKKKKKKKKKHNAETVELEKSOKNKKKKKKKKKKKKKKKKKKKK
 VKSYTV"
 BASE COUNT 2008 a 1082 c 1074 g 1212 t
 ORIGIN
 alignment_scores:
 Quality: 5739.00 Length: 1445
 Ratio: 4.477 Gaps: 12
 Percent Similarity: 88.720 Percent Identity: 78.062
 alignment_block:
 US-09-811-045A-1 x AF352051 ..
 Align seg 1/1 to: AF352051 from: 1 to: 5376
 1 MetMetGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsnH 17
 |||||||
 158 ATGATGGAAGTGAAGATCTAATATGAAAGTGCAATGCTTCAACAAC 207
 |||||||
 17 rGlyLysTyrAlaIleProThrIleAspAlaGluAlaIleGly 34
 |||||||
 208 TGGAAAATATGCAATACCACTATAGATGACAGAAAGATTAATTAATG 257
 |||||||
 34 yslsglulysPropheleuProgluluproserSerSerGlu 50

```

258 AGAAAGAAACCTCCCTCTTACAGAGAGGACATCTTCTCAGAA 307
51 GIUAsAsProIIeProIaIeUleUeCysLeuIIeCysLysAspII 67
308 GAGAGATGATCTATCCAGATGAAATTGTGTCTATCTCGAAGATAT 357
67 eMeUhrAsPaIaVaIaIIeProCysCysGlyAsnSerSerCysAsp 84
358 TATGACTGATGCTGTGTGATTCCTGCTGTGGAACAGTACTGTGATG 407
84 IuCysIIeaRgThrThrLeuLeuGluSerAspLysHisThrCysProThr 100
408 AATGATAGAACAGACATCTCGAATCAGATGACACATGTCGAGC 457
101 CyAhIeGInAsnAspVaIeSerProAsPaIaLeuIIeAlaAsnLysPhe 117
458 TGTCAATCAAAATGATGTTCTCTGATGCTTATATGCCAAATAATTTT 507
117 uATgInAlaVaIaAsnAsnPhelYsAsnGluThrGlyTyrThrLysArgL 134
508 ACCACAGGCTGTAATTAATCAAAATGAACCTGCTATACAAAAGAC 557
134 eUArgLysGInLeuProProPheLeuPheLeuVaIProProArgPro 150
558 TACGAAACAGTTACCTCTCCACACCCCAATACCACTCCGAGACCA 607
151 LeuSerGInArgAsnLeuGInProArgSerArgSerProIIeLeuArgL 167
608 CTGATTCAGAGAACCTACAACTCTGATGATGATCTCGATATCAAGACA 657
167 nGInAsPProVaIaIaPheArgTyrThrValSerProThrCysSerAsp 184
658 ACAAGATCCCTATGATGATCCAGATCATCTTCATCACT..... 697
184 hIryStrhAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPro 200
698 .....CACCA 703
201 AlapRoserIIeSerSerLeuThrSerAsnGInSerSerLeuAlaProPr 217
704 GCTCCGCTATATCTTCAATTAATCTTAATCAGTCTCTGGGCCCTCC 753
217 oValSerGIVAsnProSerSerAlaProAlaProVaIProAspIIeThrA 234
754 TGTCTTGGAATCCGCTCTCTGCTCCAGCTCTGTACTGATTAATCG 803
234 IatThrValSerIIeSerValHisSerGluYsSerAspGlyProPheArg 250
804 CAAAGATCCATATCAATTCATTCAGAAAAATCAGATGACCTTTTCGG 853
251 AspSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerGInHisSe 267
854 GATTCGATTAATAAATATTCAGCTCCAGCTCCAGCTCTGATCAGAGCATC 903
267 rLysGIVAlaSerSerIIeAlaIIeThrAlaLeuMetGluGluYsGlyV 284
904 AAAGGAACCTCTCAATTCGAATTAACCGCTTATGGAAGAGAGGCTT 952
284 alProGlyThrSerProTrpAsnSerIIePheValGlyGInSerLeuLe 300
953 A.CCAGGCGCTGTTCTTGAACCCCATCTTGTGAGCAGTCAATTAAT 1002
300 uHisGIVAlaLeuLeuProThrThrGlyProVaIArgIIeAsnAlaAla 317
1003 GCATGACAGATTGATCCCACTAGCTCCAGTAAGAAATAATACTGCTC 1052
317 rGProGlyGlyGlyArgProGlyTyrTrpGInHisSerAsnLysLeuGlyTyr 333
1053 GTCCAGGTGGTGGTGCAGCAGGTGGGAACATTCACAAACTTGGCTAT 1102
334 LeuValSerProProGInGInIIeArgArgGlyGluArgSerCysTyrAr 350

```

```

1103 CTGGTTTTCACACACAAATTAAGAGGGAGAGGAGCTGCTACAG 1152
350 gSerIIeAsnArgGlyArgHisHisSerGluArgSerGInArgThrGInS 367
1153 AAGTATAAACCGTGGGCGACACACGAGGAAAGATCAGAGAGACTCAAG 1202
367 eRProSerLeuProAlaThrProCysPheValProValProProProPro 383
1203 GCCGTCACCTACAGCAACTCCAGTCTTGTGTACCTTTCACACACTGCT 1252
384 LeuTyrProProProThrHisThrLeuProLeuProGlyAlaProPr 400
1253 TGTATCCGCTCTCCCACTACACTTCTCCCTCCGAGTGTCTCTCC 1302
400 oProGInPheSerProGInPheProSerSerGInProProThrAlaGlyT 417
1303 TCCACAGTTTCTCTCAGTTTCTCTCTGCGCCACCAACCCGCTGGGT 1352
417 yISerValProProProGlyPheProProAlaProAlaAsnIIeSerThr 433
1353 ATAGTGTCCCTCTCCAGGTTTCTCCAGCTCTCGCAATTTATCAACA 1402
434 AlAcysPheSerProGlyVaIProThrAlaHisSerAsnThrMetProTh 450
1403 CTTGGGTATCATCAGAGAGTGCAGACGCTCATTAATTCATCCCAAC 1452
450 rThrGInAlaProLeuLeuSerArgGluGluPheTyrArgGluGIn... 465
1453 AACCAAGACCAACCTTTGTCAGGGAGAAATTCATAGAGACAGCGAC 1502
466 .....Asn 466
1503 GACTAAAGAAAGAAAGAAAGAAAGTCAAGCTAGATGATTACAAAT 1552
467 AspLysGlyArgGlu..... 471
1553 GATTTTCTAAGAAATGATGAAATACAAAAGATTCMAAGAGAGCTAG 1602
472 .....SerLysPheProTyrSerGlySerSerTyrSera 483
1603 GCGCTCATTTCCAGGTCAATCTCCTATAGGTGTTCTGATTTCA 1652
483 rGSerSerTyrThrAspSerSerGInGlyLeuAlaGInHisIleHisAl 499
1653 GAAGTTATATCTATTATTAATCAAGATCTGGTCAACCGTTTCAGC 1702
499 alEuthrLeuSerProSerAlaAlaHisThrLeuAspLeuHisAsp 516
1703 TCTTATCTCGATCATTCACGCCGCTCACATTCCTGCTTATTCACGCTC 1752
516 ISProHisProProGInGluGluAlaArgSerAlaMetIleValHis 532
1753 ACCTCCATACCCCAAGAGGCGAGAGCGAAGCGCAATTACCGTTTCAC 1802
533 MeTrProAspLeuMetAspIIeAlaHisAlaArgSerArgSerProTyr 549
1803 GGTATAGTCCATGATATCAT...CGATCTAGTCAAGGTACACCCCTTA 1851
549 rArgArgTyrArgSerArgSerArgSerProProGluPheArgGlyGInS 566
1852 CAGACGCTATCATTCACGATCAAGATCTCTCAAGCCTTTAGGGGACGT 1901
566 eRProThrLysArgAsnVaIProArgGluGluYsGluArgGluTyrPhe 582
1902 CTCCTAATAAACGTAAATGTACTCAAGGGCAACAGAAAGTAATTTT 1951
583 AsnArgTyrArgGluVaIProProProTyrAspIIeLysAlaTyrTyrG 599
1952 AATGATACAGAGAAGTTCACCAACATATGACATGAAGATATTTATGG 2001
599 yArgSerValAspPheArgAspProPheGluGluYsGluArgTyrArgLut 616
2002 GAGAAAGTGTGACTTTAGAGACCATTTGAAAGAAAGACGTACGAGAA 2051

```

| | | |
|------|---|-------|
| 916 | ThrArgArgProArgSerLeu...ArgLysIleAsnTrpLeuIleAlaIar | 931 |
| 2943 | AAAGAAAGGCCAAAGAGCGCTATGTAGAAAAACAAACCCTGTAATATATAG | 29922 |
| 931 | gGluLysAsnGluATrpgGluLysArgLys...LysSerValAspLysAspPh | 947 |
| 2993 | GGAGAAAAAAGAAAAAGAAAACTGCAACAAGAAAAAGCGTAGATTAAGATT | 30422 |
| 947 | eGluSerSerMetLysIleSerLysValGluLysThrGluLeuValI | 964 |
| 3043 | TGAGTCTTCTTCAATGAAGAAATCTGCAAACTGAACTGACTGAATATAGTA | 30922 |
| 964 | ysProSerProLysArgLysMetGluLysAspValGluLysLeuGluArg | 980 |
| 3093 | AACCATACCAAAAGCGCAAAATAGAACTGATACGAAAAATGGATATAGG | 31422 |
| 981 | ThrProGluLysAspLysIleAlaSerSerThrThrProAlaLysLysI | 997 |
| 3143 | ACCCCTGAAAAGAGCAAAATT...TCTTTAAGTGGCGAGCAAAAAAAT | 31892 |
| 997 | eLysLeuAsnATrpgLutThrGluLysLysIleGluLysAsnLysAla | 10142 |
| 3190 | CAAACTCAACACAGAAACTCGGGAAGAAAAATTGGAAGTACAGAAATATAT | 32392 |
| 1014 | eThrThrLysLysLysProSerGluLysLeuGluSerThrSerSerLysIle | 10302 |
| 3240 | CAAAACCAAAAGAACCCCTCTGAAAAATTGGAGTCAACATCTAGCAAAATT | 32892 |
| 1031 | LysGluGluLysValLysGluLysAlaLysArgLysValAlaGlySerG | 10472 |
| 3290 | AAACAAGAAAAAGTCAAAAGAAAGGTCTGACAGCAAAAGAGCTGGAACGTA | 33392 |
| 1047 | uGlySerSerSerThrLeuValAspTrpLysThrSerThrSerSerThrGly | 10642 |
| 3340 | AGAGTCCACTCAACTCTGTGTGATTACACAGCTACAGAGCTCAACTGAG | 33892 |
| 1064 | LysErProValArgLysSerGluLysLysThrAspThrLysArgThrVal | 10802 |
| 3390 | GCAGTCTCTGGCGAAATCTGAGAAAAAACAGATACAAACGCAACTGTG | 34392 |
| 1081 | IleLysThrMetGluLutTrpAsnAspAsnThrAlaProAlaGluAs | 10972 |
| 3440 | ATTAAACATGAGCAAGAAATATATATATGACAAATACCGCGCCAGCGAAGA | 34892 |
| 1097 | ValIleIleMetIleGluValProGluSerLysTrpAspLysAspAsp | 11142 |
| 3490 | TGTATCATTAATGATTACAGTTCCTCCTCAATCCAAATGGGATTAAGATGCT | 35392 |
| 1114 | heGluSerGluLutGluAspValLysThrThrGluProIleGluSerVal | 11302 |
| 3540 | TTCGAATCTGAGAGAGATGTATAATCCACACAGCTATATCAAGTGTA | 35892 |
| 1131 | GlyLysProSerSerIleIleLysAsnValThrThrLysProSerAlaThr | 11472 |
| 3590 | GGAAAAACCGTAGAGTGTATTAATAAAATGTATGTAACAAAGCCATCAAAAT | 36392 |
| 1147 | raLysTrpTrpThrGluLysGluSerGluLutProGluLysLeuGluLysL | 11642 |
| 3640 | AGTCAGAGTATCTCTGAGAAAAAGAGAGAGCCATCCAGAAAAATTCGAAT | 36892 |
| 1164 | eubProLysGluAlaSerHisGluLeuMetGluHisGluLeuArgSerSer | 11802 |
| 3690 | TCACCAAGAGCGAGGCCATTAATCATATCAACATCAGAGGTTAAAAAGTTCA | 37392 |
| 1181 | LysGlySerAlaSerSerGluLysGluArgAlaLysAspArgGluHisSe | 11972 |
| 3740 | AAAAACTCTGCATCTAGTAAAAAAGGAAAAACCAAAAGATTCGGAATTATTC | 37892 |
| 1197 | rlLysSerGluLysAspAsnProAspLysArgLysSerGluAlaGluProA | 12142 |
| 3790 | AGGTGTTGGAAAAAGAGAACCTCTGAAAAAGAGAAAGAACACACTACGACG | 38392 |
| 1214 | sPLysGluSerThrValAspArgLeuSerGluGluGluHisPheLysThr | 12302 |

3840 AAAAAAGAGTAATTTGGACCGCTCGAATGAACAAAGAAATTTTAAAGCT 3889
1231 LeuSerGlnSerSerLysGluThrArgThrSerGluLysHisGluSerA 1247
3890 CTGTCTCAATCTTCCAAAGAGCGTAGACGTACGATTAACATGATTCAC 3939
1247 lArgGlySerSerAsnLysAspPheThrProGlyArgAspLysValA 1264
3940 TCGTGGCTTCCTCAAAATAAAGACTCTCACTCCCATAGACAAAAAAAC 3989
1264 sPTyTrAspSerArgAspTyrSerSerSerLysArgArgAspGluArgGly 1280
3990 ACTATGACACACAGAGAGTATTCAGTTCCAACGTAAGATGAAAGAAAT 4039
1281 GluLeuAlaArgArgLysAspSerProProArgGlyLysGluSerLeu 1297
4040 GAATTACAGACAGAAAGACCTCTCTCCGAAATAAAGATTCGCATC 4089
1297 rGlyGlnLysSerLysLeuArgGluGluArgAspLeuProLysGly 1313
4090 TGGACAGAAAATAATCAACCAAGGAGAGAGATTTGCCTAAAAAGGAA 4139
1314 .AlaGluSerLysLysSerAsnSerSerProProArgAspLysLysPro 1329
4140 CAGGAGATTCCAAAAAAGTAATTCAGTCCCTCAAGACAGCAAGAAACCT 4189
1330 HisAspHisLysAlaProTyrGluThrLysArgProGlyGluGluThr 1346
4190 CATGATCACAAAGCCACTTATGACTAAAGCCAAATGAAGAGACAA 4239
1346 sProValAspLysAsnSerGlyLysGluArgGluLysHisAlaAlaGlu 1363
4240 ACTGTAGATTAATAATCCTGTAGAGATGTGAGAAAGCATGATTTGAG 4289
1363 lArgAsnGlyLysGluSerSerGlyAlaAsnCysHisValTyrLeuTh 1379
4290 CAGGAAACATTAAGAGTCAGAGGCAATTAACCTTATTAATTAATTAAC 4339
1379 rArgGlnThrLeuProTyrArgArgSerThrLeuGluArgTyrArg 1396
4340 CCACCAAGACACACAGGTTGAAAAAGCAATTAATGAGCAATTGACAA 4389
1396 rGlaLProSerSerArgAsnProSer 1404
4390 GAGTACTGTCAAGCCTAAACCCAGT 4415
seq_name: gb_AC010321
seq_documentation_block:
LOCUS AC010321 207256 bp DNA linear PRI 18-DEC-2001
DEFINITION Homo sapiens chromosome 16 clone CTD-2540M10, complete sequence.
ACCESSION AC010321
VERSION AC010321.8 GI:17921202
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 207256)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207256)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 207256)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA

REFERENCE 4 (bases 1 to 207256)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 18, 2001 this sequence version replaced gi:15776689.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sshgc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.4.
FEATURES
source
1. 207256
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="CTD-2540M10"
BASE COUNT 61164 a 43640 c 45330 g 57122 t
ORIGIN
alignment_scores:
Quality: 4135.50 Length: 3462
Ratio: 3.443 Gaps: 23
Percent similarity: 34.691 Percent identity: 30.156
alignment_block:
US-09-811-045A-1 x AC010321/rev ..
Align seg 1/1 to reverse of: AC010321 from: 1 to: 207256
86 lLeaGThrThrLeuLeuGlnSerAspLysHisThrCysProThrCysHis 102
99095 ATTAAGACAGCAGCTCTCGAATCAGATGACACACATGCGCAGCTGCA 99046
102 sGlnAsnAspValSerProAspAlaLeuLeuAlaAsnLysPheLeuArg 119
99045 TCAAAATGATGTTCTCCGTGATGCTTAATTCCAATAAATTTTACGAC 98996
119 lN..... 119
98995 AGTAACTGTCTGTATCCATTTTATGAAAAAATCTTTTAACTGATTTA 98946
119 119
98945 ACTGTACTTCACTGAATCTGCATTAACATTTTCTGCAATTATTGCTTG 98896
120AlaValAsnAsnPhelLysAsnGluThrGlyTyrThrLysA 133
98895 GATCTGTAGCGCTGTAAATTAACCTCAAAAATGAACCTGCTATACAAAA 98846
133 rLeuArgLysGlnLeuProProPheLeuPheLeuValProProArg 149
98845 GACTACGAAACAGTTACCTCTCCACACACCCCAATATCACCTCGAGA 98796
150 ProLeuSerGlnArgAsnLeuGlnProArgSerArgSerProLeuArg 166
98795 CCACGTGATTCAGAGACACTACACCTCGATGAGATGTCGATATCAAG 98746
166 gGlnGlnAspProValValPheArgTyrThrValSerProThrCysSerA 183
98745 ACAACAAAGTCTCTTATGATTCACGTACATCACTCATCACT..... 98703
183 sPTyThrLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeu 199
98702CAC 98700
200 ProAlaProSerLysSerSerLeuThrSerAsnGlnSerSerLeuAlaPr 216
98699 CCAGCTCGGTATATATCTTCAATTAATCTATATCAAGTCTCTTCGCCCC 98650
216 oProValSerGlyAsnProSerSerAlaProAlaProValProAspLleT 233


```

|||||
98649 TCCTGTCTGGAAATCCGCTCTGCTCAGCTCCTGTACCGATATATA 98600
233 hralaThrValSerIleSerValHisSerGluYsSerAspGlyPro... 248
98599 CTGGACACAGATCCATATCATGTTCAATTCAGAAAAATCAGATGAGCCTTTT 98550
248 ..... 248
98549 CGGTAAAGCCTGTGTGTTTTTCTACTGTTAGAAACCAATGAGTCAATGAT 98500
248 ..... 248
98499 GTAGTGTGTTTGTGTGTTATCATCTTTTAAACAGCTTTCTTACTGAGCCA 98450
248 ..... 248
98449 CTGTGCTCAGTATATGTGGATGACTGTGAAGATGACTAGTCTTATTTG 98400
248 ..... 248
98399 GGGGAGAGAGAAACAACGTGATTTGATTTGGCAATGAGGATTAAGCAT 98350
248 ..... 248
98349 CTTGAGGTCAAAATACTACACAGGTGGCAACAGCCCTATTAAAGTTG 98300
248 ..... 248
98299 AATGTTCCATATACATACACAGCTCTGAACCTTAAACAGGTGGAGAAAT 98250
248 ..... 248
98249 CACTTGTGATCATATGATGATTTTGGATAGAGTAGTTTGTGATGTAAC 98200
248 ..... 248
98199 TAAGAACAGTTTAGATAGTAGAGAGAGGGGTAAAAACACAGATACT 98150
248 ..... 248
98149 AGAGAGTCAGGAATAGTAGTTGTTTGGACATAATTATATAGTAGAG 98100
248 ..... 248
98099 AATAGTAAACTCACAGCCTACCACCTTCAGAAATGAGACTTTGTCCAT 98050
248 ..... 248
98049 GTAGTTTAATGCTCATTAATCTTTAAAAAAGGTGTAG 98000
248 ..... 248
97999 AGGCCATTTTCAGACACACTGAGCAAAATCTTTAGATGAGACCTTTAC 97950
248 ..... 248
97949 ATTTACATTAGAAATCCACTGATCTGTGGAATTTATTGTGTGATA 97900
248 ..... 248
97899 GTGGAATAGAGCCAGAGAGTTCTATGGCTGAGCCAAAGATTATCTTTCA 97850
248 ..... 248
97849 GTTTAGTGTAGACTAGAGCTAAACTTAACCTTAATCTAGGGGAT 97800
248 ..... 248
97799 TTTTGTCTTCATAGCCAGACTTTGATTTTAAGTTTACTATTCTTGCAA 97750
248 ..... 248

```

```

97749 CTGATTTCTGTATTAACCTAGATATCTTTTAAAGAGGTCCTCATTTTA 97700
248 ..... 248
97699 CTTTTCACCTTACTACATATAGAGAACTTTCAAGTTTCAGCCTTCTCT 97650
248 ..... 248
97649 CACTTTCCTCAGAGTATTCATCATCTTTGACTTAACTATCTTAACCTGC 97600
248 ..... 248
97599 AGGTATGGGTAAACAGCAGCAAACTTTTACTCTTAATGTGCATTTTC 97550
249 ..... Phe..ArgAspSerAsp 254
97549 TGATAAATTTTAAATGATATATAATTTATGTTTAAAGGATTCGATA 97500
254 snlYsIleuProAlaIleuThrSerGluHisSerLysGlyAla 270
97499 ATAAATATTGGCAGCTGCAGCTCTTGATCAGACACTCAAAAGGAGACC 97450
271 SerSerIleAlaIleThrAlaIleuMetGlu... 281
97449 TCCTCAATTGCATTTACCGCTTTATGAGAGAGAGATAATTACTGG 97400
281 ..... 281
97399 TGCTTAATTTGGGATTTTAAACAGAGGTTTGTGTTCTGGCTTT 97350
281 ..... 281
97349 GTTCTAACCTCTTTTTTTTTTAACTATTCTGTATTATTTGTCG 97300
282 ..... LysGlyValProGlyThrSerProTrpAsnSerIlePhe.V 295
97299 GTTGTCTTAAAGGTTACAGAGTGCCTGTTCTGGAACCCCATCTTTCG 97250
295 alGlyIleSerIleuHisGlyIleuIleProIleThr... 308
97249 TTGACAGTCATTTATTCATGACAGTGTATCCCAACTGGTGAATA 97200
308 ..... 308
97199 GATCAGTTGGTTAGACCTTTCCCTTAAATAATTTTAACTGTA 97150
309 .....G 309
97149 TTTAATGTCAATTTTGTATTTACTACCTGCTTTTATTTTGTGTTCTAG 97100
309 LyrProValArgIleAsnAlaIleArgProGlyGlyIleArgProGlyTrp 325
97099 GTCCAGTAAAGATAATATACGTCGTCACAGTGTGTGTCACCAAGGCTGG 97050
326 Glu... 326
97049 GAACAGTGAATGATGTTTACAAATCTTCAGATGATTCCTGCAGAACT 97000
326 ..... 326
96999 AGATTAATGGAAGTCCAAACTAGCAAACTAGATAATGGAAGTCCAAAC 96950
326 ..... 326
96949 TAGCAGCGGGTCGTCCTTTATGTGAACCTTCATGTGCCAGATTA 96900
326 ..... 326
96899 CATTTTATGTAAATGTCAAAATTCAGCTTATATGAGGCTTTTTC 96850
326 ..... 326
96849 AATGAGCAAAATCTTTTCTAAATACAAATTTTAAAGTTTTTTTCTAC 96800

```

```

326 ..... 326
96799 TTTCATATCCAAACAAATGGGATCTTGACAGACTTCAGTTAATGATA 96750
326 ..... 326
96749 GATAGAAATTACTGAAAGTATTTAAATATGCGCAGTATTTTCC 96700
326 ..... 326
96699 TAGTAGTTAACTTTTTCACCTTAATTTGGATGATCAAGTTTGATT 96650
326 ..... 326
96649 TGAGTACATAAATACATAATAGATTACGCTACCAACAAAATTTT 96600
326 ..... 326
96599 TGTCTGAATCGAGTGTCTTCGCCACCTAAATACAAAGAACTTTGAT 96550
326 ..... 326
96549 TTTCCTGTGGGATATTTTATGCAATTAATAATCAGCAGACTTACTTAAT 96500
326 ..... 326
96499 GTATTTTACTATATATTAATTTCTGTTATATATCCGATGAAAT 96450
326 ..... 326
96449 GATATATTCTTAAAGTAGTAGTACTAGCAGATGATCAGTAGAAGA 96400
326 ..... 326
96399 ACATAGATTAGAAACAGTAGACTTGGTTGAATCCCATCTCTTCA 96350
326 ..... 326
96349 TTTCCTAGCTCATGACAGTGTCTGTGTTTACTCCCTAACATCAT 96300
326 ..... 326
96299 TTCCCTGTTTGATTTGGAATGACACTACCTTAACATGTAATAATGA 96250
326 ..... 326
96249 CTTCATGTGCCAGACAGACAAAAGGCTTTCTCTGTATATACAAAT 96200
326 ..... 326
96199 AAGTTCACAGATAATTATAGAGTCTAATTAATAATATATACATA 96150
326 ..... 326
96149 AAGTAATCTATTTGAGGTCTAGTTATTTGATGATACAGTTAATGTG 96100
326 ..... 326
96099 GGGGAAAGTAGTTTAAGTAATAATATATTTACTCAGCTCTGATCCG 96050
326 ..... 326
96049 AAACAAATTTAAGAACTGGGAGTAGAAAAAGTTTCAGAGTAGTTAAA 96000
326 ..... 326
95999 ATTGCTGCTTAACAGCTATGCAAGGAAAGTCAATATTAATACCAGG 95950
327 ..... His.SerAsnLysLeu 331
95949 AGAAATACTTAGCCTATTTTCTTCTCCACACAGTTCCACAAACTT 95900

```

```

332 GlyTyrLeuValSerProProGlnGlnIleArgArgGlyLysSerCys 348
|||||
95899 GGCTATCTGGTTTCTCCACCAACAATTAGAAGGGAGAGAGCTG 95850
348 STYR..... 349
|||||
95849 CTACAGGTAGGCATGCTAAATACTTGAAAAATATATCATCTATTTT 95800
349 ..... 349
95799 TATATATCTCACAAAGAAATTAAGCACAGCAGTGGCAGAGGAAG 95750
349 ..... 349
95749 GGGGTCAATTTGTTTATGTTGCTTATCCCTTAAGTCTGGGTAGTTG 95700
349 ..... 349
95699 TCCTTAAAGAGGCTAGAAAAGATAATATATATATATATATATGA 95650
349 ..... 349
95649 TGTTTTAAACATTTTATATATCTTCCAGATTAATCTTTCCTT 95600
349 ..... 349
95599 GTCGCAACAGACAGATGATTTGGCTTGCATTTACCTCTTGAAATGAA 95550
349 ..... 349
95549 GAAAGTCAGGGGTACAGACATATCTAGATACAGTTAATTAAGATTA 95500
349 ..... 349
95499 AAGTTAACAAAGCTCAGCTCCAACTGTGAACAAAGAAATTCGAACTATTCT 95450
349 ..... 349
95449 CTTCAGACCCCATCTGCCAATCATCTCCACCGCTCTTTTCCAAAG 95400
349 ..... 349
95399 AAGAGACAGAGATTGTTAGAGCTCCTGCCCTAGTGGCGGGTACAAGC 95350
349 ..... 349
95349 AATATTAATCTAAGAAATAGAAATCTTACTGTAGTCAAGGGTCCCA 95300
349 ..... 349
95299 ACCCTGGGCCACAGACCAAGTATCCATCTTGGCTTGAAGAACTGGGCA 95250
349 ..... 349
95249 ACACAGCAGAGTGAAACAGCAGCTGAGCAGATTAACACCTGAGTCC 95200
349 ..... 349
95199 ACCCTGTGTAGATCAGCAGTGGCATTAAGATCTCATAGAGCATGAAC 95150
349 ..... 349
95149 CCTGTTGAACTGGCTGCAAAAGATCTGGGTTGCACGCTCTTANGA 95100
349 ..... 349
95099 GAATCTAATGCCGTGATCATTAAGTGAACAGTTTACCCTGCGCC 95050
349 ..... 349
95049 CCGTCCATGAAAAATTTGCTTCCACAAACAGCTCCCTGTGCCAAAC 95000
349 ..... 349

```

94999 GGTGAGAACTGCTAGTCTAGTCAATAATCACACCAGGATAAGAGCC 94950
349 349
94949 TCCAAAACTTCATTAACAGAGTTTGGCCAGCAATATGGCTCATGCTGG 94900
349 349
94899 TAATTTAGAGCTTTGGAGGCTGATGGGGAGATGCTTGAGGCTGA 94850
349 349
94849 AGTCAAGATTAGCCTGGGCAACATAGCAAGACTGTCTACCAAAAA 94800
349 349
94799 AATAAAATTAACAGACATGGTGGCATGCACGTGTGCTCCAGCTACCTGG 94750
349 349
94749 GAGGCTGAGGCGAGAGATCACTTAGCTTAGGAAGTCAAGCTGCAGTC 94700
349 349
94699 AGCTATAACCATCATCTGCACCTTAGCTTAGGGGACAGAGTGAGACCT 94650
349 349
94649 GTCTCCAAAAAGAGAAATTCGGATATCTCTAGTGGCTCATGGCTAA 94600
349 349
94599 ACTTATGAAAGTTTACTATGTCTACTAGTCAGATCATTTTAAAGAGGT 94550
349 349
94549 AATAGTTATTCAATGGCTCAGTGAAGATGCTTAAATCTAAACCCCTAC 94500
349 349
94499 TCGTGTGTTGTGTACCAATTTATTAATTTAAAGCACCTCTAGTTAA 94450
349 349
94449 CAGAACTAACAGTCTGTTGGTGAGCAAAATTAGAACTTAGCCAA 94400
349 349
94399 CACTGTATTTTTACTTAGCTTTTGAGAGAAAGCTTGATATTTTCATT 94350
349 349
94349 GTTAATTATAGAATTAAACTAGAAATTTATCTGCTTTGATGATCTTT 94300
349 349
94299 ACCCACTTCCTCCCTCAAAAAAGAGGTACAGAGTTGTTTCTCTTCAT 94250
349 349
94249 TTAACCTTTTGGGGCAGGTGTGTGCTCTCACTGTAAACCCAGCACT 94200
349 349
94199 TTGTGAGATCACAGTGTGAGATTGAGAACAGCTGCCCAACATG 94150
349 349
94149 GTGAACCTGTCTCTTAATAAATACAAAAATAGTGGGCAATGGTGGC 94100
349 349

94099 ACACGTCTAGTCCAGCTACTTGGGAAGCTGAGCAGAGAAATTCGCT 94050
349 349
94049 TGAACCCAGAGGTGAGGTGCAGTGAGCCAGAAATTACACACTGTACT 94000
349 349
93999 CCAGCCTGGACGACAGAGTGAGACTCTGTCTCAAAAAATAAAAATAAAT 93950
349 349
93949 GATAAGTAACTTTTGGGAGAGTAGGTCAATAGCAGCATTAACCTGTC 93900
349 349
93899 AGTGCTTCAAAAAGACAGTTTAGAAGTAAACAGTAGTAATTAGTGAGCT 93850
349 349
93849 GATTAATCACTGTAAGCTTCAATAAAGAACTTTAATACAAATATTTC 93800
349 349
93799 GTGTAGAGAAGTCTATTAACCTTCATCATCATCAGCAGACATGGCCT 93750
349 349
93749 TAAAAATTATGTTAGAGTTATTTTACAAATAAATGTACTAATATGTTG 93700
349 349
93699 CTACACTAGTGAATAGTTGGTTCTATTCCACTGTTTAAAGTTAAAA 93650
349 349
93649 TAATTATGATAGAGAGATAAGCTTACTGTCTCAGATGACTTAGTTT 93600
350 ArgSer11LeuSerArg134 356
93599 GAAGAATTAATATCTTGAATTTATTTTCCAGAAATATAAACCGTGCGC 93550
356 rghHisSerGluArgSerGlnArgThrGlnSerProSerLeuProAla 372
93549 GACACCACACGCAAAATCACAGAGACTCAAGGCCCTGCTACTACACAGA 93500
373 ThrProCysPheValProValProProProLeuTyrProProPro 389
93499 ACTCCAGCTTTGTACTGTCTCCACACACCTTTGTATCCGCTCTCC 93450
389 OH1ThrLeuProLeuProProGlyValProProProGlnPheSerProG 406
93449 CCATACACTTCCCTCCCTCCGGGTCTCTCTCCACAGTTTCTCTC 93400
406 LnpheProSerSerGlnProProThrAlaGlyTyrSerValProProPro 422
93399 AGTTTCCTCTGGCCAGCACACACCCGCTGGGTATGCTGCTCTCCCA 93350
423 GlyPheProProAlaProAlaSerThrAlaCysPheSerProG 439
93349 GGGTTTCCTCCAGCTCTGCAATTTATCAACCTTTGGGTATACACAG 93300
439 yValProThrAlaHisSerAsnThrMetProThrThrGlnAlaProLeu 456
93299 AGTGCAGACAGCTCATTAATACATCCCAACACACAGCACCCCTT 93250
456 euserArgGluGluPheTyrArgGluGln 465
93249 TGTCCAGGGAAGATTTATAGAGACAGCAGACTAAAGAGAGTAT 93200
465 465
93199 GATTTCTAATTTGAATTTATATACACTTTTTCATTTTCTGTATTAATAC 93150

465 465
93149 ATTAATAGTGTCTGGAAAGCTTTCCAGTGTTCAGTAGATGAGA 93100
465 465
93099 AATTCACCTAAATTTAAACCAAGCCATACAGCACTCTCAGGATTT 93050
465 465
93049 CCTAGTTATAGTTTAAAGTGAACCTAATGGCAATTCATTCATTTC 93000
465 465
92999 ATTGATGATCTAATTACATTTGTCTGTAGTTTCTTAATGGTTTTC 92950
465 465
92949 TTAACCTTCACCTTTTACTTTATTTTACTAGGAAAGAAAGTCCAA 92900
465 465
92899 GCTAGATGAGTTTACAAATGATTTTGTCTAGGAATGATGAAATACAAA 92850
466 .AsnAspLysGly..... 469
 :::|||||
92849 AGATTCAAAAGAGCGTAGGGCTCATTTTCCAGTTAGTGTTTTGCAG 92800
469 469
92799 CCTTCACACAGAAATTACAAAGGACTTGAATATCCAGGAAATAGCTA 92750
469 469
92749 TGGATATAAAGACATTGCTACTAGGAATTTTCTGATTCGGTCTAGTGC 92700
469 469
92699 TATAGAAAGTAAGATGATGATTATTTCCCAATATCTGTTCATTGAATAA 92650
469 469
92649 TGCTCAGTGTAAATTTATTTGTGTAGAACTACAGTGTGGGATTTAAA 92600
469 469
92599 TTGCCCTTGTCCCTTAAAGTTATTCATATGATCAATATTTGGAAATGT 92550
469 469
92549 TTTTGAAGAAAGAAATATGAGGGAATACCAAAAACAGTTCACGAGAC 92500
469 469
92499 CTTTACCATAGGCTACAGCTCTCTGTGGAACATTAACTGCGTAATTT 92450
469 469
92449 ATGAAGTAGTAGTGAATACAGATCGCTTTATTGATAGAAAAAGTGAATTT 92400
469 469
92399 ATAAAAATTCCTGGAGTTTGTTTAAAGACTTATTTTTCAGCTTAC 92350
469 469
92349 ACTTGAATGTGTTTCTTTTAAACATGAGCTGTGGCCAGTATAGACA 92300
469 469
92299 AAAACAGACTCTACTTCAATATTCAGATAGGAAGTGAATTTAAATTT 92250

469 469
92249 AATTGACATTTTGTTTATGTAGTAGCTCATATACTACAGGTACTTT 92200
469 469
92199 AATGCACCTACCTAAATTTGTAAGACTTCACTTTTGTGTGAACAAGT 92150
469 469
92149 GATTGCTCCGATTTGTGACTTTATTCATTCACATGAGAGTCACTGTC 92100
470 Arg 471
92099 CTCCTTAATCATCTAAGATGAATTTGATATAAAGATGATGAACCGTG 92050
471 Lu..... 471
92049 AGAAAGAAATGCTAGAAATCTGAGAGTCTGTGTTTATTTGTATGTAATTT 92000
472 SerLysPheProty 476
91999 CTCACATATCACTAATGTTTGTGTTTGTGTTTGTAGCTATAATCTCCCTA 91950
476 rSerGlySerSerTySerArgSerSerTyThrAspSer.SerGlnGly 492
91949 TAGTGGTTCTCTGATTCAGAGAGTTCATATCTATCTTAATCAAGAT 91900
493 LeuAlaGlnHisIleHisAlaLeuThrLeuSerProSerAlaAlaHisIstb 509
91899 CTGGTTCAACACGTTCAACGCTTATCTCGATCATTCATTCAGCCGCTCAT 91850
509 rLeuAspLeuLeuHisAspHisProHisProGluGluAlaAla 526
91849 TCTGTTCTTATTCACGCTCACCTCCATACCCCAAGAGAGCAGAGCAA 91800
526 rGSerAlaMetIleValHisMetProAspLeuMetAspIleAlaHisAla 542
91799 GAGCCGCAATTCACGTTCCACGCTAGATCTCATGATATCAT.CGATCT 91751
543 ArgSerArgSerProProTyrArgArgTyrArgSerArgSerArgSerPr 559
91750 AGGTCAGGCTACCCCTTACAGAGCGTATCATTCACGATCAAGATTCCTC 91701
559 oProGluPheArgGlyGlnSerProThrLysArgAsnValProArgGlu 576
91700 TCAAGCGTTTAGGGGACAGCTCTCTAATAAACGTAATCTACTCAAGGG 91651
576 LuLysGluArgGluTyrPheAsnArgTyrArgGluValProProTyr 592
91650 AAACAGAACTGGAATATTTTATAGATACAGAGAACTCCACACCATAT 91601
593 AspIleLysAlaTyrTyrGlyArgSerValAspPheArgAspProPheG 609
91600 GACATGAAGCATATTTATGGAAGAGTGTGACTTTGAGACCCATTGGA 91551
609 uLysGluArgTyrArgGluTyrPgluArgLysTyrArgGluTyrPglu 626
91550 AAAAGAACGTCACCGAGATGAGAGAAATATAGAGAGGTGTGAAA 91501
626 ystYrTyLysGlyTyrAlaValGluAlaGlnProArgProSerAlaAsn 642
91500 AATATTTAAAGTTATGCTGCTGAGACAGCTAGACCTCAGCAAT 91451
643 ArgGluAspPheSerProGluArgLeuLeuProLeuAsnIleArgAsn 659
91450 AGAGAGAACTTTCTCCAGAGAGATTTTTCACCTTAACATCAGAAATTC 91401
659 rProPheThrArgGlyArgArgGluAspTyrAlaAlaGlyGlnSerHis 676
91400 TCCCTTCACAGAGCGCCAGAGAGATGATGTGTGGCAAGGATATA 91351
676 rGAsnArgAsnLeuGlyGlyAsnTyProGluLysLeuSerThrArgAsp 692


```
1165 oLysGLuAlaSerHisGluLeuMeGlnHisGluLeuArgSerSerLysG 1182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89512 CAAGACGTCGACCATGAAATCATACAAATGAGCTTAAAGTCCAAA 89463
1182 LysSerAlaSerSerGluLysGluArgAlaLysAspArgGluHisSerGly 1198
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89462 ACTGTCATCTAGTGAAGAAAGGAAACCAAGATCGAGATTATTCAGTGC 89413
1199 SerGluLysAspAsnProAspLysArgLysSerGlyAlaGlnProAspLys 1215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89412 TTGGAAGAAAGGAAACCTGAAAGAGAGAAAGCACTCCAGCAGAGAA 89363
1215 SGLSerThrValAspArgLeuSerGluGlnGlnHisPheSerThrLys 1232
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89362 AGAGAGTAAATTGGACCCCTCGAATGAAACAAACGAAATTTTAAAGCTCTGT 89313
1232 eRGLnSerSerLysGluThrArgThrSerGluLysHisGluSerValArg 1248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89312 CTCATATCTTCCAAAGAGGCTGAGAACGTCAGATTAACATGATTCACCTCGT 89263
1249 GlySerSerAsnLysAspPheThrProGlyArgAspLysLysValAspTyr 1265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89262 GCTTCTCTAAATTAAGACTCTCCTCCATAGAGACAAAGAAAACGACTA 89213
1265 rAspSerArgAspTyrSerSerSerSerLysArgArgAspGluArgLysGluL 1282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89212 TGACACCCAGAGATGATTCAGTTCACAAACGTAGAGATGAAAGATGAT 89163
1282 euAlaArgArgLysAspSerProProArgGlyLysGluSerLeuSerGly 1298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89162 TTAACAGAGCGAAAGACCTCTCTCGGATTAAGATTCGATCTGCA 89113
1299 GlnLysSerLysLeuArgGluGlnArgPleuProLysLysGly...Al 1314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89112 CAGAAAATTAACCAAGGAGAGAGAGATTGGCTTAAAGAAAGAACAGG 89063
1314 aGlnSerLysLysSerAsnSerSerProProArgAspLysLysProHisA 1331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89062 AGATTTCACAAAAGTAATTCTAGCCCTCAAGACACGAAACCTCATG 89013
1331 sPHSLysAlaProTyrGluThrLysArgProCysGluGluThrLysPro 1347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89012 ATCCAAAGCCACCTATGATGATTAACCGCCAAATGAGAGACAAATCT 88963
1348 ValAspLysAsnSerGlyLysGluArgGluLysHisAlaGluAlaArg 1364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88962 GTAGATTAATAATCCTGTGAGGATCGTGAGAACATGATTAAGACAG 88913
1364 gAsnGlyLysGluSerSerGlyAlaAsnCysHisValTyrLeuThrArg 1380
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88912 GAACATTAATAAGTCAGTGCAGTAAACCTATTATTAATTAACCCAC 88863
1381 GlnThrLeuProTyrArgArgSerTyrPleuLeuGlyArgTyrArgAla 1397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88862 CAGAGACACAGGTTGAAAGAGCAAAATTAATTAATTAATTAATTA 88813
1397 aProSerSerArgAsnProSer 1404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88812 ACTGTCAAGCCTAAACCCCACT 88791
seq_name: gb_pr:HSRBQ1
seq_documentation_block:
LOCUS HSRBQ1 3011 bp mRNA linear PRI 29-MAY-1996
DEFINITION H.sapiens RBQ-1 mRNA.
ACCESSION X85133
VERSION X85133.1 GI:728590
KEYWORDS RB protein binding protein; RBQ-1 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
```

```
REFERENCE 1 (bases 1 to 3011)
AUTHORS Sakai,Y.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-1995) Y. Sakai, Biology Div., National Cancer
Center Research Inst., Tsukiji 5-1-1, Chuo-ku, Tokyo 104, JAPAN
REFERENCE 2 (bases 1 to 3011)
AUTHORS Sakai,Y., Saijo,M., Coelho,K., Kishino,T., Nlkawa,N. and Taya,Y.
TITLE cDNA sequence and chromosomal localization of a novel human
protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene
Product Genomics 30 (1), 98-101 (1995)
MEDLINE 96129310
FEATURES
source location/Qualifiers
1..3011
/organism="Homo sapiens"
/strain="NCI-H69"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p11.2-12"
/clone="1"
/cell_line="NEC"
/cell_type="small cell lung carcinoma"
/clone_1b="lambda ZAP1"
1..2994
/gene="RBQ-1"
/product="RB protein binding protein"
/evidence="experimental"
1..2994
/gene="RBQ-1"
92..2938
/gene="RBQ-1"
/codon_start=1
/product="RB protein binding protein"
/protein_id="CAA59445.1"
/db_xref="GI:755748"
/db_xref="SPTREMBL:Q15290"
/translacion="MGIKTLNLYLGLKRALPEPEYFMEEKDPNNKGAAMLTNTGKVAL
PTIDEAYALIGKKEKPPLEPPSSSEDDPIPELCLICKDINTDAVVPICGNS
YCDICRTALIDSDEHTCPTCHONVSPALANKFLROAVNNKREKTYRLKQL
PPPPPIPPRPLIORNIQPLMRPSIROODLMIIPVSS7HPAPSSISNOSLS
APPVSGNPSAPAPPTDTATYTSVSHSKSGPPRDSNKLTPAALASBSKSTSS
IAITLAMEKGVOVPLGTPSLGSLGSLHGLIPTGGPVRIINTARGGRPOEWSNK
LGIVSPPOOIRGERSKCYRSINRGHNSERQOGPSLPATPVFVPPVPPPLVPP
PHTLPPVPPPOEPPPCGPPAGSVPPPGPPAPAMLSIPVWSSGVOTASHN
RTPTPOAPLSREEFYREORLKEEEKKSLIDEFNDAKELMEKTKIOKRRRSFS
BKSPVSGSSVSRSYTSKRSRSGTSSRSRSRSRSRSRSRSPVPRGRKSR
NYSRSRSGHYHRSRSRSPYRRYHRSRSPAPRQSPNNKNVPOGETREYENRNR
EVPPEYDMKAYIGRSYDFRDEPEKERTREWEKRTKMTKGTAGAGQPRPSARE
NPSPERFLPLNIRNSPFTGRREDYVGGSHRNIGSNYPKLSARDGHNOKDKWS
KEKESNAGDGKGNKHKRRKRKGESEGLINPELETSSRSRSPTEGVEBKDLSL
FVLPSRHDATPVDEPMDEISITPFSVSEKDRERDPRKAGDKTKRKMDGSAVSKK
LTVKPAKGPOEKVDVDDVLDLNLQIKKPKRRRLRLTLNHLPLRRKKKSLPEPEK
LTLNODKTRNKTSQDKSEBGLFQRCQIRKANN"
```

alignment_scores:
Quality: 3938.00 Length: 978
Ratio: 4.633 Gaps: 8
Percent Similarity: 86.912 Percent Identity: 78.937

alignment_block:

US-09-811-045A-1 x HSRBQ1 ..

Align seg 1/1 to: HSRBQ1 from: 1 to: 3011

```
1 MetetGluValLysAspProAsnMetLysGluAlaMetLeuThrAsnTh 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
158 ATGATGAGAGGAAAGATCCCTAATATGAAAGGTGCAATGCTTACCACAC 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 rGlyLysTyrAlaIleProThrIleAspAlaGluAlaTyrAlaIleGlyL 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```


[illegible]

82 yAspGluCysIleArGThrThrLeuLeuGluSerAspLysHisThrCys 98
 |||||.....
 8491 GCGATGAATGTATTAAGACAGACACTTCTGGAGTCAGATGACATACATCATGT 8442
 99 ProThrCysHisGlnAsnAspValSerProAspAlaLeuIleAlaAsnLys 115
 |||||.....
 8441 CCAACATGTCATCAAAATGATGTTCTCTGATGCTTAGTTAGTGCACAA 8392
 115 sPheLeuArGlnAlaValAsnAsnPhelysAsnGluThrGlyThrL 132
 |||||.....
 8391 GGTTTTACAGACAGGCTGTAAATTAACCTTAAATAACTGACTACACAA 8342
 132 yArGLeuArGlyGlnLeuProProPheLeuPheLeuValProPro 148
 |||||.....
 8341 AGAGACTGGGAAAACAGGCTACTCCGTCGCCACCCCCACTACTCTCCACA 8292
 149 ArGProLeuSerGlnArGAsnLeuGlnProArGSerArGSerProIleLe 165
 |||||.....
 8291 ACTGACATCATTCAGACAGAACCTGCAGCCTCCTATGAATACTCTCACATC 8242
 165 uArGLeuGlnAspProValValPheArGThrThrValSerProThrCys 182
 |||||.....
 8241 TTAGACAGCAGGATCCTGTGAAGATTCCAATGACA..... 8208
 182 eArSpThrLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArG 198
 |||||.....TCGCTCAGCCTCAGCT 8190
 8207TCGCTCAGCCTCAGCT 8190
 199 LeuProAlaProSerLleSerSerLeuThrSerAsnGlnSerSerLeuAl 215
 |||||.....
 8189 CACCCAACTCCTGTACTACCTCATAGCTCAAAATTCATCTTCCTCGGC 8140
 215 aProProValSerGlyAsnProSerSerLarProAlaProValProAspI 232
 |||||.....
 8139 TCCCTCTGTGCTGGAAACCACTCTCTGCCAGCTCCCGTACCGGACA 8090
 232 LeThrAlaThrValSerLleSerValHisSerGluLysSerAspGlyPro 248
 |||||.....
 8089 CCACGTGCAGAGATCTGTATCAGACCTTCAGAAAATCAGATGATGC 8040
 249 PheArGAspSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerG 265
 |||||.....
 8039 TTTCCGGGAATCAGAAAACCAATTAATACAGCGCCGCCCTTACATCAGA 7990
 265 uHisSerLysGlyAlaSerSerLleAlaIleThrAlaLeuMetGluL 282
 |||||.....
 7989 ACATTCATAAGGAGCCTCTTCATTCGTGTACTGCTCTATGGAAGAA 7940
 282 ySGLyAlProGlyThrSerProThrAsnSerLlePheValGlyGlnSe 298
 |||||.....
 7939 ACGCTGG.CCAGGTGGCGCTCTTGAACCTCCACCTTCGTTGGGACAGTC 7891
 298 rLeuLeuHisGlyGlnLeuIleProThrThrGlyProValArGLeuAsn 315
 |||||.....
 7890 ACTACTCCATGAACAGTCTATCCCCACAACCTGGTCCAGTAAGAAATG 7841
 315 lAlaArGProGlyGlyArGArGProGlyLysLysHisSerAsnLysLeu 331
 |||||.....
 7840 CTGCTCATCCAAAGTGTGTCAACAGGTGGGACATTCCAACCAAGAT 7791
 332 GlyTyrLeuValSerProGlnGlnIleArGArGlyGlyLysSerC 348
 |||||.....
 7790 GGGAGCTAGTTTCTCCACACAGCAAAATTAAGAGTGAAGAAAGAGCTG 7741
 348 sTyrArGSerLleAsnArGlyArGHisHisSerGluArGSerGlnArG 365
 |||||.....
 7740 TTAACGAGGAATTAACGTGGGAGACATAAGTAAACATCACAGAG 7691
 365 hGlnSerProSerLeuProAlaThrProCysPheValProValProPro 381
 |||||.....
 7690 GTCDAAGGCCATCAGCAGCAACCTCAGTCTTGTGCTATTCACACA 7641
 382 ProProLeuTyrProProProHisThrLeuProLeuProProGlyVal 398

|||||.....
 7640 CCTCTTTCCACACCGTCTCTCCACATATGCTTCCTCCGCCAGGTGT 7591
 398 lProProProGlnPheSerProGlnPheProSerSerGlnProProThr 415
 |||||.....
 7590 ACCTGCCCCACAGGTTCCCTCGGGTCTCTCTCCGCCAGCTCCCAACAG 7541
 415 lAlGlyTyrSerValProProProGlyPheProProAlaProAlaAsnIle 431
 |||||.....
 7540 CAGGATATATGTCTCTCCAGGATTTCCACAGCTCTCTCCCAATATA 7491
 432 SerThrAlaCysPheSerProGlyValProThrAlaHisSerAsnThr 448
 |||||.....
 7490 TCAACAGCTGTCTTTCACACAGGTGTCCACAGTCCCAATCAATCAAT 7441
 448 cProThrThrGlnAlaProLeuLeuSerArGLeuGluThrArGLeuG 465
 |||||.....
 7440 GCCCACAACCAAGCACCCTTTTGTCCAGGAGAAATTCATAGAGAGC 7391
 465 lAsnAspLysGlyArGLeuSerLysPheProTyrSerGlySerTyr 481
 |||||.....
 7390 AAAACGACTAAAGAGAGTCTAAATTTCCCTATATGTGGTCAATGAT 7341
 482 SerArGSerSerTyrThrAspSerSerGlnGlyLeuAlaGlnHisIle 498
 |||||.....
 7340 TCAAGAGTTCAATACACCTGACCAAGTCAAGGTCTGGCTCAGACATTC 7291
 498 lAlaLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHis 514
 |||||.....
 7290 ACGCTCTTACTCAGTCTCTGAGCTGCTCAGCCTCTCATCTTCTTCAC 7241
 515 AspHisProHisProProGluGluAlaGluAlaArGSerAlaMetLys 531
 |||||.....
 7240 GATCATCCCATCTCTCAGAGAGCGAGAGCAAGTCTGCAATGATGT 7191
 7190 TCACAGTCCAGATCTCATGATATGCCCATCTCAGTCAAGGTCCCTTC 7142
 548 rGlyTyrArGArGlyTyrArGSerArGSerArGSerProProGluPheArGly 564
 |||||.....
 7141 CCTATGAGAGATATCTCAGCGTCCAGATCTCTCCACATTTAAGGGA 7092
 565 GlnSerProThrLysArGAsnValProArGLeuGluLysGluArGly 581
 |||||.....
 7091 CAGTCTCCCACTAACTCGATGACTCTCGAGAA..... 7059
 581 rPheAsnArGlyTyrGluValProProProTyrAspLleLysAlaTyr 598
 |||||.....
 7059 7059
 598 ySGLyArGSerValAspPheArGAspProPheGluLysGluArGlyArG 614
 |||||.....
 7058GAGAAAGACCTTCCAG 7041
 615 GluThrGluArGlyTyrArGLeuTyrGluLysTyrLysGlyTyr 631
 |||||.....
 7040 GAATGGAGAGGAATATCACTACAGTGTATGAGAAAT...AGGCCAAAGGC 6994
 6993 AAGCATTAGGAACACAGCCTAGACCTTAGGCCAATAGAGAGGACTTTCT 6944
 648 ProGluArGLeuLeuProLeuAsnIleArGAsnSerProPheThrArG 664
 |||||.....
 6943 TCAGAGAGACTTACTCTCTGATATCAGAAATTCACCTTCAAGAGAG 6894
 664 yArGArGLeuAspTyrAlaAlaGlyGlnSerHisArGAsnArGLeu 681
 |||||.....
 6893 GCACAGAGAACTATGCTGTGACAAAGTCAATGAATAGAAAGTCTAG 6844
 681 LysGlyAsnTyrProGluLysLeuSerThrArGAspSerHisAsnAlaLys 697
 |||||.....

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 861: contig of 861 bp in length
*
* 862 961: gap of 100 bp
*
* 962 1566: contig of 605 bp in length
*
* 1567 1666: gap of 100 bp
*
* 1667 2720: contig of 1054 bp in length
*
* 2721 2820: gap of 100 bp
*
* 2821 4833: contig of 2013 bp in length
*
* 4834 4933: gap of 100 bp
*
* 4934 7443: contig of 2510 bp in length
*
* 7444 7543: gap of 100 bp
*
* 7544 10612: contig of 3069 bp in length
*
* 10613 10712: gap of 100 bp
*
* 10713 13264: contig of 2552 bp in length
*
* 13265 13364: gap of 100 bp
*
* 13365 15293: contig of 1929 bp in length
*
* 15294 15393: gap of 100 bp
*
* 15394 20002: contig of 4609 bp in length
*
* 20003 20102: gap of 100 bp
*
* 20103 26889: contig of 6787 bp in length
*
* 26890 26989: gap of 100 bp
*
* 26990 33039: contig of 6050 bp in length
*
* 33040 33139: gap of 100 bp
*
* 33140 38303: contig of 5164 bp in length
*
* 38304 38403: gap of 100 bp
*
* 38404 46409: contig of 8006 bp in length
*
* 46410 46509: gap of 100 bp
*
* 46510 54712: contig of 8203 bp in length
*
* 54713 54812: gap of 100 bp
*
* 54813 70402: contig of 15990 bp in length
*
* 70403 70502: gap of 100 bp
*
* 70503 86873: contig of 16371 bp in length
*
* 86874 86973: gap of 100 bp
*
* 86974 101977: contig of 15004 bp in length
*
* 101978 102077: gap of 100 bp
*
* 102078 128046: contig of 2569 bp in length
*
* 128047 128146: gap of 100 bp
*
* 128147 161179: contig of 33033 bp in length
*
* 161180 161279: gap of 100 bp
*
* 161280 188228: contig of 26549 bp in length
*
* 188229 188328: gap of 100 bp
*
* 188329 222549: contig of 34221 bp in length.
*
*
* location/Qualifiers
*
* 1. 222549
*
* /organism="Mus musculus"
* /db_xref="taxon:10090"
* /clone="RP23-184F6"
* /clone_lib="RPCI-23 Female Mouse BAC"
*
* misc_feature
*
* 1. 861
*
* /note="assembly_fragment
* clone_end:SP6
* vector_side:left"
*
* misc_feature
*
* 962. 1566
*
* /note="assembly_fragment"
*
* misc_feature
*
* 1667. 2720
*
* /note="assembly_fragment"
*
* misc_feature
*
* 2821. 4833
*
* /note="assembly_fragment"
*
* misc_feature
*
* 4934. 7443
*
* /note="assembly_fragment"
*
* misc_feature
*
* 7544. 10612
*
* /note="assembly_fragment"
*
* misc_feature
*
* 10713. 13264
*
* /note="assembly_fragment"
*
* misc_feature
*
* 13365. 15293

```

```
misc_feature /note="assembly_fragment"  
15394..20002  
/note="assembly_fragment"  
misc_feature 20103..26889  
/note="assembly_fragment"  
misc_feature 26990..33039  
/note="assembly_fragment"  
misc_feature 33140..36303  
/note="assembly_fragment"  
misc_feature 38404..46409  
/note="assembly_fragment"  
misc_feature 46510..54712  
/note="assembly_fragment"  
misc_feature 54813..70402  
/note="assembly_fragment"  
misc_feature 70503..86873  
/note="assembly_fragment"  
misc_feature 86974..101977  
/note="assembly_fragment"  
misc_feature 102078..128046  
/note="assembly_fragment"  
misc_feature 128147..161179  
/note="assembly_fragment"  
misc_feature 161280..188228  
/note="assembly_fragment"  
misc_feature 188329..222549  
/note="assembly_fragment"  
BASE COUNT 63391 a 45670 c 46003 g 65450 t 2035 others  
ORIGIN
```

```
alignment_scores:  
Quality: 2576.50 Length: 725  
Ratio: 4.404 Gaps: 9  
Percent Similarity: 80.690 Percent Identity: 72.414
```

alignment block:
US-09-811-045a-1 x AC099701 ..

Align seg 1/1 to: AC099701 from: 1 to: 222549

```
1 MetMetGluValIlyAspProAsnMetIySgLYAlaMetLeuThrAsn 17  
148226 ATGATGGAAGTGAAGATCCTAATACATGAAGAGTGCAATGCTTCAAAAAC 148275  
17 rGlyIySTYrAlaIleProThrIleAspAlaGluAlaTYrAlaIleGlyL 34  
148276 TGGGCAATATGCATACCGACTTAAATGCGAGGCTATGCAATTGGGA 148325  
34 yLylSgLIuIySProPhoPheLeuProGIuIuPro.....SerSerSer 48  
148326 AGAAAGAGAAACACCCCTTCTTACAGAGAGAACCATCATCATCATCATCT 148375  
49 SerGIuGIaSPasProIleProAlaGluLeuLeuCySerIleCySly 65  
148376 TCAGAGAAAGTTATCTCTGTCCAGACGAGCTTGTGCTCATCTGCGCA 148425  
65 sasPIleMetThrAspAlaValIleProCySgIyAsnSerSerC 82  
148426 AGACACTATGATGATGCTGCTCATCCCTGCTGTGGAACAGATTACT 148475  
82 yaSPuLIuIySIIeaRgThrThLeuLeuGluSerAspIySHstHrCyS 98  
148476 GTGATGAATGTATAAGACGACCTTGTGAGTCATGAACATTACATGT 148525  
99 ProThrCySHISgIlnAsnAspValSerProAspAlaLeuIleAlaAsnIy 115  
148526 CCAACATGTCATCAAAATGATGTTCTCCGATGCTTAACTTGCACAA 148575  
115 sPheLeuATrGlnAlaValAsnAsnPhelYsAsnGIuThrGlyTYrThL 132  
148576 GGTTTACGACAGCGTGTATAACTTTAAATAACAACTGACTCCACAA 148625
```

```
132 yaRgLeuATrGlySgIlnLeuProPhoPheLeuPheLeuValProPro 148  
148626 AGAGCTCGGAAAACAGTCACTCCCTCCCAACCCCATCTACTCCACCA 148675  
149 ArgProLeuSerGIlnArgAsnLeuGlnProArgSerArgSerProIleLe 165  
148676 AGTGCATCATTCAGCAGAACCTGACGCTCCATGAATATCCACATC 148725  
165 uARgGIlnAsnAspProValValPheATrGlyThrValSerProThrCyS 182  
148726 AAGACGACAGATCCTGTGAAGGTTCCAGTCA..... 148759  
182 eRAsPThrIySThrAlaGlySerCySerAspSerGIyThrLeuSerArg 198  
148760 .....TCGTTCTCAAGT 148771  
199 LeuProAlaProSerIleSerSerLeuThrSerAsnGlnSerLeuAl 215  
148772 CACCCAACTCCCTGTGATACCTCATAGCTTCMAATTTATCTCTCGCGC 148821  
215 aProProValSerGIyAsnProSerSerAlaProAlaProValProAspI 232  
148822 TCTTCTGTGCTGTGAACCCATCTTCTGCCGACGTCAGATACGTATA 148871  
232 IeThrAlaThrValSerIleSerValHISerGIuIySerAspGIyPro 248  
148872 CAACCTCAAGAGATCTATATATCAGACCATTCAGAAAATCAGATGARCC 148921  
249 PheArgAspSerAspAsnIySLeuLeuProAlaAlaIleAlaIleThrSergI 265  
148922 TTTCGGGATTCAGATTAACAAATATATATCAGCTGCCGCTTACATCAGA 148971  
265 uHISerIySgIyAlaSerSerIleAlaIleThrAlaLeuMetGluGlu 282  
148972 ACATTTAAAGGAAGCCCTTCAATTCGTGTCTTATTCAGAA 149021  
282 ySgIyAlProGIyThrSerProTrpAsnSerIlePheValGlyGlnSe 298  
149022 AGTGTAG.CCAGGTGCCAGTCTTGAACCTTCACCTTTGTTGGACAGTTC 149070  
298 rLeuLeuHISgIyGlnLeuIleProThrThrGlyProValArgIleAsn 315  
149071 ACTATATACATGAACAGTTATATCCCACTGCTTCATTAAGATAAAG 149120  
315 lAlaIaRgProGIyGlyArgProGIyTrpGluHISerAsnIySLeu 331  
149121 CTGCTCATCCAAAGTGTGTGCAACGAGTTGGAGACATCCCAACAAGCTT 149170  
332 GlyTYrLeuValSerProProGlnGlnIleArgArgGIyGluArgSerCy 348  
149171 GGGAACTGATGTTCTCCACACGACGAAATTAAGAGTGAAGAAAGACTCTG 149220  
348 sTYrARgSerIleAsnArgIyArGhISHISerGIuARgSerGlnARg 365  
149221 TTACCGAGGAATTAACGTGGGCAAGACTACAGTGAACATCACAGAGCA 149270  
365 hGlnSerProSerLeuProAlaThrProCySPheValProValProPro 381  
149271 GTCAAGGCCACGACGACGACGACCACTCCATCTTGTGCTATTTCCACCA 149320  
382 ProProLeuTyRProProProHISThrLeuProLeuProProGIyVa 398  
149321 CCTCTTATGCACTGTCTCCACATATGCTTCTTCTTCTCCAGGTGT 149370  
398 lProProProGlnPheSerProGlnPheProSerSerGlnProProTha 415  
149371 ACCTGCCACGAGGTTCTCCCTGTTGTTCTCTCCTAGCAGACTCCACAG 149420  
415 lAGlyTYrSerValProProProGIyPheProProAlaProAlaAsnIle 431  
149421 CAGGATATATGTGCCCTCCGAGGATTTCCACACACTCTCTGCAATATA 149470  
432 SerThrAlaCySPheSerProGIyValProThrAlaHISerAsnThrMe 448
```

```

||||||| ||| ||||| |||||||||||||||||
149471 TCACACACCTGGTTCATGCAGAGATTCACACACTGCTCATTCATAATACCAT 149520
448 tProthThGlnAlaProLeuLeuSerArgGluLupheYrArgGluG 465
:|||||||||||||||||||||||||||||||||
149521 CCCCAACAACACACACCTCTTTGTCAGAGGAAGAAATTCATATGAGAGC 149570
465 InAsnApLysGlyArgGluSerLysPheProYrSerGlySerSerYr 481
||||| :|||||
149571 AAAAACACATRAAGAAAGAGCTAAATTTCCCTATAGTGGTCAATCGAT 149620
482 SerArgSerSerYrThrAspSer.SerGlnGlyLeuAlaGlnHisLeu 498
||||| :|||||
149621 TCAGAAGTTCATACACTGACTCAAGCTCAAGTCTGGCTCAACACATTC 149670
498 ISAlaLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHis 514
||||| :|||||
149671 AGGCTCTTACTAGTCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 149720
515 AspHisProHisProProGluGluAlaGlnAlaArgSerAlaMetIleVa 531
||||| :|||||
149721 GATCATCCCATCTCCAGAAAGAGCAGAGCAAGATCTGCATGATTTGT 149770
531 LHisMetProAspLeuMetAspIleAlaHisAlaArgSerArgSerProp 548
||||| :|||||
149771 TCACAGTCCAGATCTCATGATATCAC.CGATCTAAGGTCAAGGTCCCTCC 149819
548 roYrArgArgYrArgSerArgSerArgSerProGluPheArgGly 564
||||| :|||||
149820 CCTATAGACATATCACTCACGCTCCAGATCTCTCCAGCTTTAAGGGGA 149869
565 GlnSerProThrLysArgAsnValProArgGluGluLysGluArgGluTy 581
||||| :|||||
149870 CAGCTCCCATTAACCTGATGTAACCTCAAGAA..... 149902
581 rPheAsnArgTyrArgGluValProProProTyrAspIleLysAlaTyrT 598
149902 ..... 149902
598 YrGlyArgSerValAspPheArgAspProPheGluLysGluArgTyrArg 614
149903 .....GAGAAAGAACCTACCAG 149920
615 GlnTrpGluArgLysTyrArgGluTrpTyrGluLysTyrTyr.LysGluT 631
149921 GAATGGGAGAGGAATACTACCAAGGTATGAGAAAT...AGCAAAGGCG 149967
631 YrAlaValGlyAlaGlnProArgProSerAlaAsnArgGluAspPhe.Se 647
149968 AAGCAGATAGGAACAACAGCTAGACCTAGACCTAGCCAAATAGAGAGTTTCTTC 150017
647 rProGluArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArg 664
150018 A...GAGAGACTTTCCTCTTAATATCAGAAATTCATCTTCACAGAG 150064
664 LysArgArgGluAspTyrAlaAlaGlnLysSerHisArgAsnArgAsnLeu 680
150065 GGCACAGAGAGAGACTATGCTGTCGACAAAGTCATAGAAATAGAAATCTA 150114
681 GlyLysAsnTyrProGluLysLeuSerThrArgAspSerHisAsnAlaLys 697
150115 GGTGGCAACTATCCAGTAAAGCTTTCAACAAGAGACAGTCACAATGCAAA 150164
697 sAspAsnProLysSerLysGluLysGluSerGlnAsnValProGluAspG 714
150165 AGATATTCCTCAAAATCGAAAGAGAAAGATCAGAAATGTTCTCGACAGA 150214
714 LysGlyAsnLysHis 719
150215 GCAAAAGCAACAGAAAT 150231
seq_name: gb_pat:AR048215

```

```

seq_documentation_block:
LOCUS AR048215 1800 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 11 from patent US 5821070.
ACCESSION AR048215
VERSION AR048215.1 GI:5970558
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1800)
AUTHORS Lee W.-H. and Shan B.
TITLE Antibodies reactive with retinoblastoma binding proteins and
METHODS of using same
JOURNAL Patent: US 5821070-A 11 13-OCT-1998;
FEATURES Location/Qualifiers
source 1..1800
BASE COUNT 747 a 333 c 372 g 348 t
ORIGIN

alignment_scores:
Quality: 1838.50 Length: 490
Ratio: 4.113 Gaps: 5
Percent Similarity: 91.224 Percent Identity: 75.510

alignment_block:
US-09-811-045A-1 x AR048215 ..
Align seg 1/1 to: AR048215 from: 1 to: 1800

919 ProArgSerLeu...ArgLysIleAsnTyrLeuIleAlaArgGluLysAs 934
11 CCAAGAAAGCCTAATGAGAAAAACAAACCACTTGATATTAAGGGAGAAAAA 60
934 ngLysArgGluLysArgLys..LysSerValAspLysAspPheGluSerSe 950
61 AGAAAGAAAGAAACTGAAAGAAAGGCGTGAATAGATTGAGTCTTC 110
950 rSerMetLysIleSerLysValGlnGlyThrGluIleValLysProSer 967
111 TTCATATGAAATCTCGAAACATGAGAGAGTACGAAATAGTGAACATCAC 160
967 roLysArgLysMetGluGlyAspValGluLysLeuGluArgThrProGlu 983
161 CAAAGCCGAAATGGAACCTGATCTGAAAAAATGATGAGACCCCTGAA 210
984 LysAspLysIleAlaSerSerThrThrProAlaLysLysIleLysLeuAs 1000
211 AAGACCAAAATTT...TCTTAAAGTGGCGCAGCCCAAAAATCAAACTCA 257
1000 nArgGluThrGlyLysLysIleGlyAsnAlaGluAsnAlaSerThrT 1017
258 CAGAGAAACTGGAAGAAATTTGGAAGTACAGAAATATATCAAAACAA 307
1017 ysgLupProSerGluLysLeuGluSerThrSerSerLysIleLysGln 1033
308 AAGAACCCTCTGAAAAATTTGGAGTCAACATCTACCAAAATGAAACAGA 357
1034 LysValLysGlyLysAlaLysArgLysValAlaGlySerGluGlySerSe 1050
358 AAAGTCAAAAGGAAAGGTCAGACGAAAGTGACTGCACTGAAGATCTCAG 407
1050 rSerThrLeuValAspYrThrSerThrSerSerThrGlyLysSerPro 1067
408 CTCACACTGTGGTGAATTAACCAAGTACAGCTCAACTGAGGCGAGTCTG 457
1067 alArgLysSerGluGluLysThrAspThrLysArgThrValIleLysThr 1083
458 TGGGAAATCTGAAGAAAAACAGATCAAGCAAGCAAGTGTATTAAGC 507
1084 MetGluGluTyrAsnAsnAspAsnThrAlaProAlaGluAspValIleI 1100
|||||

```

508 AAGGAGAAATATATATATGACATACCGCCGACGTAAGATGTATCAT 557
1100 eettilegivalProgluSerlystrpaspaspaphegluSerG 1117
558 TAGATTCAGGTCTCCATTCACAAATGGATTAAGATGACTTGAATCTG 607
1117 luaglulaspValylThrThrThrProilegluSerValgluysPro 1133
608 AAGAAATATGTATTAATCCACACACCTATATCAAGTGTAGAAAACT 657
1134 SerSerleilelysasValThrThrLysProSerAlaThrAlaLysTy 1150
658 GCTAGGTATTAATAAATGTAGTACAAAGCATCAATATATAGTCAAGTA 707
1150 rThrGlulysgluSerGlulnProglulysleuGlulnLysleuProLysG 1167
708 TCCTGGAGAAAGAAAGTCCATCCAGAAATTCAGAAATTCACCAAG 757
1167 luAlaSerHisgluMetGlnHisgluLeuArgSerSerlysglySer 1183
758 ACGTGACCATGAAATCATACATCAGTAAAGTTCAAAAACTCT 807
1184 AlaSerSerGlulysgluValAlaLysAspArggluHisSerlyserG 1200
808 GCATCTAGTGAATAAAGGAAACCAAGATCGAGATTATTCAGTGTGGA 857
1200 uLysAspAspProAspLysArgLysSerGlyAlaGlnProAspLysGluS 1217
858 AAGGGAACCCCGAAAGAGAAAGACACTCCAGCCAGAGAAAGAGA 907
1217 eThrValAspArgleuSerGlulnGluHisPheLysThrleuSerGln 1233
908 GATATTGGACCTCTGAAATGAAACAGAAATTTTAAAGTCTGCTCAA 957
1234 SerSerlysgluThrArgThrSerGluLysHisgluSerValArglyse 1250
958 TCTTCCAAAGAGCGTAGAACGTCAATTAACATGATTCACGTGCTTC 1007
1250 rSerAnLysAspPheThrProGluArgAspLysLysValAspTyrAspS 1267
1008 CTCCAATTAAGACTCTCTCCATGAGACACAAAAAACTGACTATGACA 1057
1267 eArgAspTyrSerSerSerlysarGArgAspGluArggluLeuAla 1283
1058 CCAGACAGATTCATCACTCCAAACGTAGAGATCAAAAGATGAATTAACA 1107
1284 ArgArgLysAspSerProArgGlyLysgluSerleuSerGlyLnu 1300
1108 AGACGAAAGACTCTCTCTCGAATTAAGATTGTGATCTGACAGAA 1157
1300 sSerLysleuArggluLnuArgAspleuProLysLysGly...AlaGluS 1316
1158 AATATAACCAAGGAGAGAGAGATTTGCCYTAAGAAAGAAAGAGAGATT 1207
1316 eLysLysSerAspSerSerProArgAspLysLysProHisAspHis 1332
1208 CCAAAAAAAGTATTTCTAGTCCCTCAAGACAGACAAAAAACCCTCATGATCAC 1257
1333 LysAlaProTyrGluThrLysArgProCysgluGluThrLysProValas 1349
1258 AAGCCACTTATGATTAACGCGCAAAATGAAGACAAAAATCTGTAGA 1307
1349 pLysAsnSerGlyLysgluArggluLysHisAlaAlaGluAlaArgAsnG 1366
1308 TAAAAATCTTGTAGAGATCGTAGAAGCATGTATTAGAAGCAAGAGAAC 1357
1366 LysLysLysSerSerGlyLysAlaAsnCysHisValTyrleuThrArgLnu 1382
1358 ATTAAGAAGTCACTGCAATTAACCTATTATTAATCAACCAACAGAGAG 1407
1382 rLeuProTyrArgArgSerTyrleuLeuGlyArgTyrPargArgAlaProS 1399
1408 ACACAGGTGTAAGAAAGAGCAAAATTTACTGGCAAAATTGACAAAGATCTACTGT 1457

1399 eSerArgAsnProSer 1404
1458 CAGGCTTAACCCCACT 1474
seq_name: gb_htg:AC095130
seq documentation block:
LOCUS AC095130 153479 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-8F21, *** SEQUENCING IN PROGRESS ***
68 unordered pieces.
ACCESSION AC095130
VERSION AC095130.2 GI:17942059
KEYWORDS HTG, HTGS, PHASEL.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 153479)
REFERENCE
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbata,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Franz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabab,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogunh,M., Okwouon,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Ruiz,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostrati,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,K.,
Stone,H., Sulton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Unpublished
Direct Submission
2 (bases 1 to 153479)
Worley,K.C.
Direct Submission
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15625684.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

```
----- Project Information
Center project name: GCQM
Center clone name: CH230-8F21
----- Summary Statistics
Assembly program: Phrap; version 0.990329F1st call to
findPhrapList
Consensus quality: 121430 bases at least Q40
Consensus quality: 129666 bases at least Q30
Consensus quality: 136708 bases at least Q20
Estimated insert size: 119515; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.bjsg.bcm.tmc.edu/docs/genbank-draft-data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
6858: contig of 6858 bp in length
6859
6959: gap of unknown length
12717: contig of 5759 bp in length
12718
12817: gap of unknown length
12818
16979: contig of 4162 bp in length
16980
17079: gap of unknown length
17080
21262: contig of 4183 bp in length
21263
21362: gap of unknown length
21363
25724: contig of 4362 bp in length
25824
25825: gap of unknown length
29022: contig of 3198 bp in length
29023
29123: gap of unknown length
31270: contig of 2148 bp in length
31271
31370: gap of unknown length
31371
35314: contig of 3944 bp in length
35315
35414: gap of unknown length
35415
37893: contig of 2479 bp in length
37894
37993: gap of unknown length
41088
41089: contig of 3095 bp in length
41188
41189: gap of unknown length
43791: contig of 2603 bp in length
43792
43891: gap of unknown length
43892
46961: contig of 3070 bp in length
46962
47061: gap of unknown length
50215: contig of 3154 bp in length
50216
50315: gap of unknown length
50316
52419: contig of 2104 bp in length
52420
52519: gap of unknown length
55806: contig of 3287 bp in length
55807
55906: gap of unknown length
55907
59295: contig of 3389 bp in length
59296
59395: gap of unknown length
59396
61798: contig of 2403 bp in length
61799
61898: gap of unknown length
61899
64243: contig of 2345 bp in length
64244
64343: gap of unknown length
67007: contig of 2664 bp in length
67008
67107: gap of unknown length
68873: contig of 1766 bp in length
68874
68974: gap of unknown length
71449: contig of 2476 bp in length
71450
71549: gap of unknown length
71550
74329: contig of 2780 bp in length
74330
74429: gap of unknown length
74430
76070: contig of 1641 bp in length
76071
76170: gap of unknown length
76171
78425: contig of 2255 bp in length
78426
78525: gap of unknown length
80194: contig of 1669 bp in length
80195
80294: gap of unknown length
```

```
*
80295
81542: contig of 1248 bp in length
81543
81642: gap of unknown length
81643
83636: contig of 1994 bp in length
83637
83736: gap of unknown length
83737
86415: contig of 2679 bp in length
86416
86515: gap of unknown length
86516
88344: contig of 1829 bp in length
88345
88444: gap of unknown length
88445
90189: contig of 1745 bp in length
90190
90289: gap of unknown length
90290
92121: contig of 1832 bp in length
92122
92221: gap of unknown length
92222
93934: contig of 1713 bp in length
93935
94034: gap of unknown length
94036
96058: contig of 2024 bp in length
96059
96158: gap of unknown length
96159
97881: contig of 1723 bp in length
97882
97981: gap of unknown length
97982
99035: contig of 1074 bp in length
99056
99155: gap of unknown length
99156
100698: contig of 1543 bp in length
100699
100798: gap of unknown length
100799
102147: contig of 1349 bp in length
102148
102247: gap of unknown length
102248
105291: contig of 3044 bp in length
105292
105391: gap of unknown length
105392
107562: contig of 2171 bp in length
107563
107662: gap of unknown length
107663
108956: contig of 1294 bp in length
108957
109056: gap of unknown length
109057
110501: contig of 1445 bp in length
110502
110601: gap of unknown length
110602
112705: contig of 2104 bp in length
112706
112805: gap of unknown length
112806
114186: contig of 1381 bp in length
114187
114286: gap of unknown length
114287
116172: contig of 1886 bp in length
116173
116272: gap of unknown length
116273
117323: contig of 1051 bp in length
117324
117423: gap of unknown length
117424
119313: contig of 1890 bp in length
119314
119413: gap of unknown length
119414
120518: contig of 1105 bp in length
120519
120619: gap of unknown length
122209: contig of 1591 bp in length
122210
122309: gap of unknown length
122310
123432: contig of 1113 bp in length
123433
123522: gap of unknown length
123523
125754: contig of 2232 bp in length
125755
125854: gap of unknown length
125855
127342: contig of 1488 bp in length
127343
127442: gap of unknown length
127443
129015: contig of 1573 bp in length
129016
129115: gap of unknown length
129116
130755: contig of 1640 bp in length
130756
130855: gap of unknown length
130856
132296: contig of 1441 bp in length
132297
132396: gap of unknown length
132397
133904: contig of 1508 bp in length
133905
134004: gap of unknown length
134005
135616: contig of 1612 bp in length
135617
135716: gap of unknown length
```

```
alignment_scores:
Quality: 1777.00      Length: 408
Ratio: 4.664          Gaps: 5
Percent Similarity: 93.382  Percent Identity: 89.216
```

alignment_block:

US-09-811-045A-1 x AC095130 ..

Align seg 1/1 to: AC095130 from: 1 to: 153479

```

659 SerProphethrArglyArgarglyAspArglyAlaIaGlyGlnSerH 675
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
107681 AGTACTTATCAAGAGGTTCTCAGAGAGACTATGTTCTGGCAAGTGA 107730
675 sArgasnArgasnLeuGlyGlyAsnTyrProGlnLysLeuSerThrArga 692
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
107731 TGGAAATAGAAATCTAGGTGGCACTATCCAGAAAAAGATTTCAGACAGAG 107780
692 sPserHisAsnAlaLysAspAsnProLysSerLysGlnLysGlnSerGlu 708
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
107781 ATAGTCACAAATCCAAAAGATATCCAAAGTCCAAAAGAGAGAGAGTGA 107830
709 AsnValProGlyAspGlyLysGlnLysAsnLysHisLysArgLysArg 725
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
107831 AATGTTCCAGGAGCGCAAGAGAAACAAAGCATAGAAACACAGGAAAG 107880
725 gArgasnGlnLysGlyGlyGlnLysSerLysLeuAsnProGlnL 742
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
107881 AAGA ..... AAGGGGAGAGAGGTGAGAGCTTCGAAACCCAGAGC 107921
742 euLeuGlnThrSerArgLysCysArgGlySerSerGlyLysAspGlnThr 758
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
107922 TACTGGAAATCTTAGAAATCCAGGAGATCCTCAGGAGATCATGAAACA 107971
759 LysThrAspThrLeuPheValLeuProSerArgAspAspAlaThrPro 775
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
107972 AAGACAGATGCACTGTTGTTCTCCAGACAGAGAGATGCTACACTGT 108021
775 lArgAspGlnProMetAspAlaGlnSerLysThrPheLysSerValSer 792
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
108022 TGGGATGAAACCAATGATGAGAGATCATCTCAATCAATCAGATATCG 108071
792 sPlysAspLysArgGlnLysAspLysProLysValLysSerAspLysThr 808
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
108072 AAAAAACAAAGAGAGAAAGATAAACCAAAAGTAAAAAGTACAGAGACC 108121
809 LysArgLysSerAspGlySerAlaThrAlaLysAspAsnValLeuLys 825
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
108122 AAACGGAAAAAGATGGTGTCTCTACAGCCCAAGAAAAAGTTTAA 108171
825 sProSerLysGlyProGlnGlnLysValAspGlnLysAspArgLysSerP 842
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
108172 ACCTCTAAAGAGACCCCAAGACAGATAGATGAGACCTGAAAAATCTC 108221
842 rArgSerGlnProProLeuLysLysAlaLysGlnLysAlaThrLysIle 858
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
108222 CTGGGTGTAACCTCTCTCAAAAAAGAGAGAGGCTACAAAGGTC 108271
859 AspSerValLysProSerSerSerSerGlnLysAspGlnLysValThrG 875
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
108272 GACTCTGTAACCT...TCTCATCTCAGAAAGATGAGAGATGCTCCTGG 108318
875 yThrProArgLysAlaHisSerLysSerAlaLysAsp...ThrArgArgGln 891
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
108319 AACCCCTAGAAAAAGCCCACTCCAAATCTCAAAAAGAACACAGAGGCAA 108368
892 SerGlnProArgThrArgArgSerLysArgThrValProLysThrSerSe 908
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
108369 ACCCACTCAAGAGAGAGGCAAAAAAGAGACTGTTCCAAAAGCAACAG 108418
908 rGlnLysSerGlnProValArgThrArgArgProArgSerLysLeuArgLys 925
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
108419 TCAGAAAAACCCAGCCAGTAGAGAGAGAGAGCCCAAGAACCTCAGAGAAAA 108468
925 lAsnTyrLeuIleAlaArgGlnLysAsnGlnLysGlnLysArgLys...L 941
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
108469 TAAACTACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 108518
941 ySerSerValAspLysAspPheGlnUserSerSerMetLysLysSerLysVal 957
||||| ..... ||||||| ..... ||||||| ..... ||||||| .....
108519 AAGAGTGA.....GATTTTGAAGTCTCTTCAAGAAAAATCTCTAAAGTG 108562
958 GlnGlyThrGlnIleValLysProSerProLysArgLysMetGlnLys 974

```

```

108563 GAGGAGACGAAATAGTGAACCATCACCAGAAACCGAAATAGAGGTGA 108612
974 pValGlnLysLeuGlnArgThrProGlnLysAspLysIleAlaSerSerT 991
108613 TGTGAAAAGCTGGAAGAGACCCCAAGAAAGAGCAAGATGCATATCA 108662
991 hThrProAlaLysLysLysLysLeuAsnArgGlnThrGlnLysLysIle 1007
108663 TTACCCACCCCAAAAAATTAACATCAATAGAGAACTGAAAAAAATTT 108712
1008 GlyAsnAlaGlnAsnAlaSerThrThrLysGlnProSerGlnLysLeu 1024
108713 GGAATGCGAAAAATCCATTATACAAAAGAACCTATGAAAAACTGGA 108762
1024 UserThrSerSerLysLysLysGlnGlnLysValLysGlyLysAlaLys 1041
108763 GTCACATCTAACAATAATCAACAGAGAAAAAGTCAAGAGAGAGCCAGAC 108812
1041 rGlnValAlaGlnSerGlnGlnLysSerSerThrLeuValAspLysThr 1057
108813 GGAAATAGCTGGAGACAGAGAGCCCAAGCTCCAGCTTGATGACTACACC 108862
1058 SerThrSerThrGlyGly 1064
108863 CAGGTATGAGAGCCCGGTGGG 108883

seq_name: gb_pr:AK026954
seq_documentation_block:
LOCUS AK026954 2320 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ23301 fls, clone HEP11120.
ACCESSION AK026954
VERSION AK026954.1 GI:10439935
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:hep2 cDNA to mRNA, clone_1lb:HEP
clone:HEP11120.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (sites)
Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Ohtani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
JOURNAL
2 (bases 1 to 2320)
unpublished (2000)
REFERENCE
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
AUTHORS
Direct Submission
TITLE
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@elms.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Departent of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
Source
1..2320
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP11120"
/cell_line="HEP2"
/cell_type="hepatoma"
/clone_1lb="HEP"
/note="Cloning vector pME18SF13"
3..1889
/note="unnamed protein product"
CDS

```


PROGRESS ***, in unordered pieces.

AL670951
AL670951.4 GI:18643842
HTG: HTGS_PHASE1; HTGS_ACTIVIFIN; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Bates, K.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On Feb 10, 2002 this sequence version replaced g1:18491458.
----- genome center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM6N18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 203030 bases at least Q40
Consensus quality: 203337 bases at least Q30
Consensus quality: 203532 bases at least Q20
Insert size: 203737; sum-of-ctrls
Insert size: 208368; 7.5% error; agarose-ff
Quality coverage: 13.20x in Q20 bases; sum-of-ctrls Quality coverage: 13.09x in Q20 bases; agarose-ff

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source location/qualifiers
1..204937
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-6N18"
/clone_lib="RPCI-23"
1..26114
/note="assembly_fragment:04643
clone_end:17
vector_side:left"
26215..35512
/note="assembly_fragment:05542
fragment_chain:1"
35613..64166
/note="assembly_fragment:04831
fragment_chain:1"
64267..69715
/note="assembly_fragment:05492
fragment_chain:1"
69816..101415
/note="assembly_fragment:01175
fragment_chain:2"
101516..109750
/note="assembly_fragment:02972
fragment_chain:2"
109851..121829
/note="assembly_fragment:01319
fragment_chain:3"
121930..131947
/note="assembly_fragment:06136
fragment_chain:3"
132048..139628
/note="assembly_fragment:03401
fragment_chain:3"

misc_feature 139729..149923
/note="assembly_fragment:02273
fragment_chain:3"
150024..165181
/note="assembly_fragment:06020
fragment_chain:3"
165282..202703
/note="assembly_fragment:03330
fragment_chain:3"
202804..204937
/note="assembly_fragment:05292
fragment_chain:3"
clone_end:SP6
vector_side:right"
BASE COUNT 52741 a 47028 c 47279 g 56686 t 1203 others
ORIGIN

alignment_scores:
Quality: 1585.50 Length: 503
Ratio: 4.129 Gaps: 10
Percent Similarity: 76.342 Percent Identity: 66.203

alignment_block:
US-09-811-045a-1 x AL670951/rev ..

Align seg 1/1 to reverse of: AL670951 from: 1 to: 204937

237 SerIleSerValHisSerGluYsSerAspGlyProPheArgAspSerAs 253
||||| ::|||
23664 AGATATTAAATCATTCAGAAAATCAGATGACCCCTTGCGGATTCAG 23615
253 PAsnIlySLeuLeuProAlaAlaAlaLeuThrSerGluHisSerIlyG 270
|||||
23614 TAACAAATTATTAACAGCTGCACCCCTTACATCAGAACATTCAAAGG 23565
270 IAsnSerIleAlaIleThrAlaLeuMetGluGluYsGlyValProGly 286
|||||
23564 CCCTTCATTTGCTGTACTGCTCTTATGGAAGAAAGTG.TGGCCAGGC 23516
287 ThrSerProTrpAsnSerIlePheVal.GlyIleSerLeuLeuHisGly 303
:::|||||
23515 GCCAGTCCTTGAACATCCACCTTTGTGGGACAGTATTATTCACAGAAC 23466
303 ILeuLeuProThrThrGlyProValArgIleAsnIleAlaAlaProGly 319
||:::|||||
23465 AGTTTATCCCACTGACAGTGCAGTAAGATAATGCTGCTCCAGGT 23416
320 GlyIlyArgProGlyTrpGluHisSerAsnIlySLeuGlyTyrLeuValSe 336
|||||
23415 GGTCGTCACAGAGTTGGGAACTTCACACAGCATGGGTACTAGTTG 23366
336 rProProGlnGlnIleArgArgIlyGluArgSerCysTyrArgSerIleA 353
|||||
23365 CCCACACAGAAAATTAGAAAGACAAAGACAGCTGTACAGAGGATAA 23316
353 snArgIlyArgHisIleSerGluArgSerGlnArgThrGlnSerProSer 369
:::|||||
23315 AACGCTGGCAACACTACAGACGACATCAAGAGAGTCAAGGCCGCTCA 23266
370 LeuProAlaThrProCysPheValProValProProProLeuLeuTrp 386
|||||
23265 CGACCCAGACATCAGCTTTTGGCTGTCCACACCTCTCTATGCAACC 23216
386 oProProProHisThrLeuProLeuProProGlyValProProGlnP 403
|||||
23215 T.....CTTCCTCTCTCTAGTATACCTCCCCACAGT 23181
403 hSerProGlnPheProSerSerGlnProProThrAlaGlyTyrSerVal 419
|||
23180 TTCTCTCTCAGTTTCTCTGCTAGCTCCACACAGGATTAAGTGGCC 23131
420 ProProProGlyPheProProAlaProAlaAsnIleSerThrAlaCysPh 436

| | | | |
|-------|---|-------|--|
| 23130 | CTCTCTCCAGGGAGTCCCAACAGCTCTCTGCCAATATATCAACAGCTGGGT | 23081 | |
| 436 | eSerProGlyValProThrAlaHisSerAsnThiMetProThrTrgIa | 453 | |
| 23080 | TTCCGACATAGTGCAGACTGCCATTCAAAATCAATACCTTATACACACAG | 23031 | |
| 453 | IaProLeuLeuSerArgLysLupheTyArgGlyLysAlaAspLysGly | 469 | |
| 23030 | CAC TTCCTTGTGCCAAGGAAAGATTTCTTTAGAGACCAACACTACTAAAG | 22981 | |
| 470 | ArgLysSerLysPheProTySerGlySerSerTySerArgSerSerTy | 486 | |
| 22980 | GAGCGCTAAATTTCCCTACAGTGGCTCATCGTATTCACAAATTCATA | 22931 | |
| 466 | rThraspSerSerGlnIleuLeuAlaGlnHisIleHisIleLeuThreu | 502 | |
| 22930 | CAC TACAC TCAAGTCAAGGTCTGGCTCAACACAGTCAACACTTACCTTA | 22881 | |
| 503 | SerProSerAlaAlaHisThrLeuAspLeuLeuHisAspHisProHisPr | 519 | |
| 22880 | AGTTCTTCAGACGCTCACACTCTGCTCTTCTTCACAGATCATCCCATCC | 22831 | |
| 519 | OProGluGluAlaGluAlaArgSerAlaMetIleValHisMetProAspL | 536 | |
| 22830 | TCCACAGAGGCGGACGAGGCAAGATCTGCATGACCATTCATGGTCCAGATC | 22781 | |
| 536 | eMetAspIleAla | 540 | |
| 22780 | TCATGGATATTCACAATCTAGTCAAGTCCCTCCCTTGATCTCATGGA | 22731 | |
| 541 | HisAlaArgSerArgSerProProTyArgArgTyArgSerArg | 555 | |
| 22730 | TATCAACATCTAGTCTCAAGGCCCCCTCCCTATACATCACTACACTCATG | 22681 | |
| 555 | GSerArgSerProProGluLupheArgGlyLysSerProThrLysArgAsnV | 572 | |
| 22680 | GTCACGATCTCTCCACGATTTTAAAGAACAGTCTCCACTAATACATCATG | 22631 | |
| 572 | AlProArgGluGluLysGluArgGluTyrrPheAsnArgTyrrArgGluVal | 588 | |
| 22630 | TCCCTCAAGA | 22620 | |
| 589 | ProProProTyrrAspIleAlaIleTyrrGlyArgSerValAspPheAr | 605 | |
| 22620 | ProProProTyrrAspIleAlaIleTyrrGlyArgSerValAspPheAr | 22620 | |
| 605 | GAspProPheGluLysGluArgTyrrArgGluTrpGluArgLysTyrrArgG | 622 | |
| 22619 | GAGACAGACGCTTACAGGAATGGGAGAGGAAGCACACAGG | 22580 | |
| 622 | IuTrpTyrrGluLysTyrrTyrrLysGlyTyrrAlaValAlaGluProAr | 638 | |
| 22579 | GCTGTATGAGAGTAAAGGCCAAAGGCAAGCAGTAAAGACATCTTAG | 22533 | |
| 638 | GProSerAlaAsnArgGluAspPheSerProGluTrgLeuLeuProLeuA | 655 | |
| 22532 | ACCCTTAGGCCAACAGAGAGCTTTTCTTCAGAGAAATCTTACTCTTA | 22483 | |
| 655 | snIleArgAsnSerProPheThrArgGlyArgArgGluAspTyrrAlaAla | 671 | |
| 22482 | ATATCATTAACCCACCTTCCCAAGGAGGCGACAGAGAGACTATGTGCT | 22433 | |
| 672 | GlyGlnSerLysArgAsnArgAsnLeuGlyLysAsnTyrrProGluLysIle | 688 | |
| 22432 | GGGCAAAAGTCGTAAATATAAGGAGCAACTTACCCGAAAAGT | 22389 | |
| 688 | uSerThrArgAspSerHisAsnAlaLysAspAsnProLysSerSer | 702 | |
| 22388 | TTCAACACAGACACAGTCCAAATAGCAAAAGATATATCCAAAGACATAACTCA | 22339 | |
| 703 | LysGluLysGluSerGluAsnValProGlyAspGlyLysGlyAsnLysHis | 719 | |

| seq_name | gb_pr | HSMB02662 |
|---------------------------|---|-------------------------------------|
| 22338 | AAAGGAGGAAGAAATGAGAAATCTTCAGAGACAGACAAAGGAAACAGCA | 22289 |
| 719 s | 719 | |
| 22288 | T | 22288 |
| seq_name: gb_pr:HSMB02662 | | |
| LOCUS | HSMB02662 | 2026 bp mRNA Linear PRI 18-JUN-2000 |
| DEFINITION | Homo sapiens mRNA; CDNA DKFP761B2423 (from clone DKFP761B2423). | |
| ACCESSION | AL359564 | |
| VERSION | AL359564.1 | GI:8655620 |
| KEYWORDS | human. | |
| ORGANISM | Homo sapiens | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | |
| AUTHORS | 1 (bases 1 to 2026) Bloecher, H., Boecher, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S. | |
| TITLE | Direct Submission | |
| JOURNAL | Submitted (15-JUN-2000) MIPS, Am Klopferplatz 18a, D-82152 Martinsried, GERMANY | |
| COMMENT | Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; Sequenced by GBF (National Research Centre for Biotechnology Ltd., Brunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFP761B2423) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heuberweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/ . Location/Qualifiers | |
| FEATURES | 1..2026 | |
| source | /organism="Homo sapiens" /db_xref="taxon:9606" /clone="DKFP761B2423" /tissue_type="amygdala" /clone_lib="761 (synonym: hamy2). Vector pSport1; host DH10B; sites NotI + SalI" /dev_stage="adult" 1..1592 /gene="DKFP761B2423" /gene="DKFP761B2423" /note="similarity to p53-associated cellular protein PACT (Mus musculus)" /codon_start=3 /product="hypothetical protein" /protein_id="CAB94869.1" /db_xref="GI:8655621" translation="TIVDYTSSTSGSGSPVAKSEKDTKRTVITKIMEYNNNDTAPAE EDVILIOVPOSKMDKDFESEEDVKSTOPTISSVGKPAVINKVSTPSNIVKPE ESESEKTKQFTKTVSHETIIIOHYKSSNSASSEKGRKDDYVLEKEENKRSNS QPEKESNIDRLNEGNGEFSLSQSKREATSPKSHSTRASSKKDPTPRNKKCTDDYD YSSSRDRDEKNELTRRKSPSRNKSQASQGNKPREEDLKKGGGDSKSSSPSRD KKPDHDKATYETLRKPRNETKSVDRNPKCDRKHVLEAKNNKSSGNKLLYLINPEYQ VEKDIQGOIDKSTYVKPKPOLSHSRSLSDLTREDEAKFEPDYNESDSESNVKEE ESSGNISKEDLKDIYERAKKESLDPAAYOVGIISSNOSHSPSPSPSRSHSPSGSVTESS HSSNASSAEQDSKK KNKDEKKEKDDKQVNSVTV" | |
| gene | | |
| CDS | | |
| alignment_scores: | | |
| Quality: | 1353.50 | Length: 355 |
| Ratio: | 4.190 | Gaps: 2 |
| Percent Similarity: | 90.986 | Percent Identity: 74.648 |

RAATVHGLDMDIIDLGGHPRITDAIHDODLLKRLGDSILLNVMYTKGKONVNIIL
DTEKFNHMT"
BASE COUNT 271 a 151 c 154 g 164 t
ORIGIN

alignment_scores:
Quality: 1046.00 Length: 247
Ratio: 4.628 Gaps: 1
Percent Similarity: 91.498 Percent Identity: 80.162

alignment_block:
US-09-811-045a-1 x BC015318 ..

Align seg 1/1 to: BC015318 from: 1 to: 740

```

483 ArgSerSerThrThrAspSer.SerGlnGluLeuAlaGlnHisLeuHisA 499
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 AGAAGTCAATATCTTATCTTAATCAAGATGCTGTTCAACAGCTCACG 52
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
499 laeuthrleuSerProSerAlaAlaHisThrleuAspLeuHisAsp 515
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 CTCTTATCTCGATCATTCACGCCGCTCACATCTCTCTCTATTCACGGT 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
516 HisProHisProGluGluAlaGluAlaArgSerAlaMetIleValH 532
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
103 CACCTCCATCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 152
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
532 smetProAspLeuMetAspIleAlaHisAlaArgSerArgSerPro 549
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
153 CGGTCTAGATCTCATGATGATATCAT.CGATCTAGTCAAGGTCAACCCCTT 201
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
549 YrArgArgTyrArgSerArgSerArgSerProGluPheArgGlyGln 565
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 ACAGAGCCCTTCATTCACGATTCAGATCTCTCAAGCTTTAGGGGACAG 251
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
566 SerProThrIysArgAsnValProArgGluGluGluGluGluGluGlu 582
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
252 TCTCTTAATTAACGTAATGTACTCTCAAGGGGAGAGAGAGAGAGAGAG 301
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
582 eAsnArgTyrArgGluValProProProTyrAspIleLysAlaTyrTyrG 599
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
302 TAAATGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 351
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
599 YrArgSerValAspPheArgAspProPheGluGluGluGluGluGlu 615
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
352 GGAGAAGGTTGACTTATAGACACCATTTGAAAAAGAGCGTACCGAGAA 401
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
616 TrpGluArgTyrArgGluGluTyrGluGluGluGluGluGluGluGlu 632
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 TGGAGAGAGAAAAATATAGAGAGTGTATGAAAAAATTTAAAGGTATGC 451
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
632 aValGlyAlaGlnProArgProSerAlaAsnArgGluAspPheSerPro 649
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 TGTGTGAGACACAGCTAGACCTTACGCAAAATAGAGAGAGAGAGAG 501
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
649 luArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArgGlyArg 665
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
502 AGAGATTGTCGACTTAACATCAGAGATTCTCCCTTCAAGAGAGCGCG 551
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
666 ArgGluAspTyrAlaAlaGlyGlnSerHisArgAsnArgAsnLeuGly 682
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
552 AGAGAGAGAGACTATGTGTGGCAAAAGTCATAGAACTCGAAATAGGAG 601
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
682 yAsnTyrProGluGluGluSerThrArgAspSerHisAsnAlaLysAsp 699
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
602 CAACATATCCAGAAAGCTTTCAGCAAGAGATGTCACAATTCAGAGAGATA 651
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
699 snProLysSerLysGluGluSerGluAsnValProGlyAspGlyLys 715
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
652 ATACAAAGTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATA 701
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
716 GlyAsnLysHisLysLysHisArgLysArgArgAsnGlu 728

```

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
702 GCAAAATACCATTAAGAACACAGAAAAAAGAAAAA 740

seq_name: gb_htg:AC095130

seq_documentation_block:

LOCUS AC095130 153479 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-8F21, *** SEQUENCING IN PROGRESS ***,
68 unordered pieces.
ACCESSION AC095130
VERSION AC095130.2 GI:17942059
KEYWORDS HTG: HTGS, PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 153479)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbarta,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,D.,
Bowie,S., Brivea,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkelt,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Day-Carroll,L., Dederich,D.A., Delanaye,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Donthiaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homsil,F., Howard,S., Huber,J., Hulky,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J.,
Kovar,C., Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegad,H., Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapue,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mauniney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,G., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newson,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokewo,S.,
Ogub,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peetey,J., Perez,L., Peters,L., Pichers,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojupokan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Stanley,H.,
Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,K.,
Stone,H., Sutton,J., Taylor,C., Taylor,T., Telford,B., Thomas,R.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalton,C.,
Wall,R., Wang,S., Ward-Moore,S., Warren,A., Wleciyk,R., Wooden,S.,
Watlington,S., Williams,G., Williamson,A., Wleciyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Direct Submission
Unpublished
2 (bases 1 to 153479)
Morley,K.C.
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15625684.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GCOM

```
Center clone name: CH230-8F21
----- Summary Statistics -----
Assembly Program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 121430 bases at least Q40
Consensus quality: 129669 bases at least Q30
Consensus quality: 136708 bases at least Q20
Estimated insert size: 119515; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 1.6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 6858: contig of 6858 bp in length
6859: gap of unknown length
6959 12717: contig of 5759 bp in length
12718 12817: gap of unknown length
12818 16979: contig of 4162 bp in length
16980 17079: gap of unknown length
17080 21262: contig of 4183 bp in length
21263 21362: gap of unknown length
21363 25724: contig of 4362 bp in length
25725 25824: gap of unknown length
25825 29022: contig of 3198 bp in length
29023 29122: gap of unknown length
29123 31270: contig of 2148 bp in length
31271 31370: gap of unknown length
31371 35314: contig of 3944 bp in length
35315 37893: contig of 2479 bp in length
37894 37993: gap of unknown length
41088: contig of 3095 bp in length
41089 41188: gap of unknown length
41189 43791: contig of 2603 bp in length
43792 46961: contig of 3070 bp in length
46962 47061: gap of unknown length
47062 50215: contig of 3154 bp in length
50216 50315: gap of unknown length
50316 52419: contig of 2104 bp in length
52420 52519: gap of unknown length
52520 55806: contig of 3287 bp in length
55807 55906: gap of unknown length
55907 59295: contig of 3389 bp in length
59296 59395: gap of unknown length
59396 61798: contig of 2403 bp in length
61799 61898: gap of unknown length
61899 64243: contig of 2345 bp in length
64244 64344: gap of unknown length
64345 67007: contig of 2664 bp in length
67008 67107: gap of unknown length
67108 68873: contig of 1766 bp in length
68874 68974: gap of unknown length
68975 71449: contig of 2476 bp in length
71450 71549: gap of unknown length
71550 74329: contig of 2780 bp in length
74330 74429: gap of unknown length
74430 76070: contig of 1641 bp in length
76071 76170: gap of unknown length
76171 78425: contig of 2255 bp in length
78426 80194: gap of unknown length
80195 80294: gap of unknown length
80295 81542: contig of 1248 bp in length
81543 81642: gap of unknown length
```

```
* 81643 83636: contig of 1994 bp in length
* 83637 83736: gap of unknown length
* 83737 86416: contig of 2679 bp in length
* 86416 86516: gap of unknown length
* 86516 88344: contig of 1829 bp in length
* 88345 88445: gap of unknown length
* 88445 90189: contig of 1745 bp in length
* 90190 90289: gap of unknown length
* 90290 92121: contig of 1832 bp in length
* 92122 92222: gap of unknown length
* 92222 93934: contig of 1713 bp in length
* 93935 94035: gap of unknown length
* 94035 96158: contig of 2024 bp in length
* 96159 97881: gap of unknown length
* 97882 97981: contig of 1723 bp in length
* 97982 99055: gap of unknown length
* 99056 99156: gap of unknown length
* 99156 100699: contig of 1543 bp in length
* 100699 100799: gap of unknown length
* 100799 102147: contig of 1349 bp in length
* 102148 102247: gap of unknown length
* 102248 105291: contig of 3044 bp in length
* 105292 105392: gap of unknown length
* 105392 107562: contig of 2171 bp in length
* 107563 107662: gap of unknown length
* 107663 108957: contig of 1294 bp in length
* 108957 109056: gap of unknown length
* 109057 110501: contig of 1445 bp in length
* 110502 112705: gap of unknown length
* 112706 112805: contig of 2104 bp in length
* 112806 114187: gap of unknown length
* 114187 114286: contig of 1381 bp in length
* 114287 116172: gap of unknown length
* 116173 116272: contig of 1886 bp in length
* 116273 117323: gap of unknown length
* 117324 117423: contig of 1051 bp in length
* 117424 119313: gap of unknown length
* 119314 119413: contig of 1890 bp in length
* 119414 120518: gap of unknown length
* 120519 120618: contig of 1105 bp in length
* 120619 122209: gap of unknown length
* 122210 122309: contig of 1591 bp in length
* 122310 123422: gap of unknown length
* 123423 123522: contig of 1113 bp in length
* 123523 125755: gap of unknown length
* 125756 125854: contig of 2232 bp in length
* 125855 127342: gap of unknown length
* 127343 127442: contig of 1488 bp in length
* 127443 129015: gap of unknown length
* 129016 129115: contig of 1573 bp in length
* 129116 130755: gap of unknown length
* 130756 130855: contig of 1640 bp in length
* 130856 132296: gap of unknown length
* 132297 132396: contig of 1441 bp in length
* 132397 133904: gap of unknown length
* 133905 134004: contig of 1508 bp in length
* 134005 135616: gap of unknown length
* 135617 135716: contig of 1612 bp in length
* 135717 135716: gap of unknown length
```

alignment_scores:

```
Quality: 907.00 Length: 201
Ratio: 4.903 Gaps: 1
Percent Similarity: 92.040 Percent Identity: 87.065
```

alignment_block:

```
US-09-811-045A-1 x AC095130/rev ..
```

```
Align seg 1/1 to reverse of: AC095130 from: 1 to: 153479
```

```
472 SerLysPheProTyrSerGlySerTyrSerArgSerSerTyrThrAs 488
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```

109663 TCTAAATCTCCCTATAGTGGTATGCTATTCAGAGTTCATACACTTA 109614
488 pser.SerGinglyLeuAlaGlnHisIleHisAlaLeuThrLeuSerPro 504
109613 CTCAAAGTCAAGTGTGGTCAACAGCTTCAGCTCTTACTCTCGGCTT 109564
505 SerAlaAlaHisThrLeuAspLeuHisAspHisProHisProProG1 521
109563 TCAGCCGCTCAGACTCTCGCTCTATTCACGCTCACCCCATACCCAGG 109514
521 uGluAlaGluAlaArgSerAlaMetIleValHisMetProAspLeuMet 538
109513 AGAGCGAGAGGCAAGAGCCCAATTACCGTTCAAGCTCCAGATCTCTAG 109464
538 spLeuAlaHisAlaArgSerArgSerProProTyrArgArgTyrArgSer 554
109463 ATACCAT.CGATCTAGTCAAGGTCACCTCCCTATACAGCATACCACTCA 109415
555 ArgSerArgSerProProGluPheArgGlyGlnSerProThrLysArg 571
109414 CGGTCCGATCTCTCTCAAGCATTTAGCGGACAGTCTCCACTAAACGCAA 109365
571 nValProArgGluGluLysGluArgGluTyrPheAsnArgTyrArgGlu 588
109364 TGTACTCTCAAGAGAAACAGAGCGTGATTTTATATAGATACAGAGAG 109315
588 alProProProTyrAspIleLysAlaTyrTyrGlyArgSerValAspPhe 604
109314 TTCACCCCTTATGATATCAAAACCTATATATGCGGAGGTGCTGACTTT 109265
605 ArgAspProPheGluLysGluArgTyrArgGluTyrPheLysTyrArg 621
109264 AGAGACCCATTGTAGAAAGACGCTACCGGAAATGGAGAGAAATACCG 109215
621 gGluTyrTyrGluLysTyrTyrLysGlyTyrAlaValGlyAlaGlnPro 638
109214 AGACTGTATGAGAACTACAAAGGTACGCTGAGGAGCCCAAGCCTTA 109165
538 rGPProSerAlaAsnArgGluAspPheSerProGluArgLeuLeuPro 654
109164 GACCTTCAGCCCAACAGAGAGACTTTTCTCCAGAGAGACTTACCTCTT 109115
655 AsnIleArgAsnSerProPheThrArgGlyArgArgGluAspTyrAla 671
109114 AATATCAAGAAATTCACCTTACAAAGAGCGCAGAGAAAGACTATGTTGC 109065
671 a 671
109064 T 109064
seq_name: gb_hhg:AL450323
seq_documentation_block: 110015 bp DNA linear HTG 10-JUL-2001
LOCUS AL450323
DEFINITION Mus musculus chromosome 2 clone RP23-349016, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL450323
VERSION AL450323.4 GI:12539773
KEYWORDS HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 110015)
REFERENCE
AUTHORS Pearce,A.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
on Jan 26, 2001 this sequence version replaced gi:11493349.
COMMENT
----- Genome Center
Center: UK Medical Research Council

```

```

Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouseghar.mrc.ac.uk
----- Project Information
Center project name: BM349016
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: pGAP4; version 108752; 100% of reads
Chemistry: dye-terminator Big Dye; 99% of reads
Chemistry: dye-primer Big Dye; 0% of reads
Consensus quality: 109929 bases at least Q40
Consensus quality: 109962 bases at least Q30
Consensus quality: 110005 bases at least Q20
Insert size: 119290; sum-of-contigs
Quality coverage: 16.70x in Q20 bases; sum-of-contigs quality
coverage: 15.58x in Q20 bases; agarose-tp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..110015
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-349016"
/clone_1lb="RPCT-23"
1..110015
/note="assembly_fragment:01877
clone_end:SP6
vector_side:left
vector_end:T7
vector_side:right"
BASE COUNT 27835 a 27870 c 26415 g 27895 t
ORIGIN
alignment_scores:
Quality: 773.00 Length: 279
Ratio: 3.944 Gaps: 8
Percent Similarity: 70.251 Percent Identity: 60.573
alignment_block:
US-09-811-045A-1.x AL450323/rev ..
Align seg 1/1 to reverse of: AL450323 from: 1 to: 110015
460 GluPheTyrArgGluGlnAsnAspLysGlyArgGluSerLysPheProTyr 476
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110015 GAATTCCTTTCAGAGCAACACACTCTTAAGAGAGAGCGTCTAATTTCCCTA 109966
476 rSerGlySerSerTyrSerArgSerSerTyrThrAspSer.SerGingly 492
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109965 CAGTGGGTGATCGATATTCACAAATTTCTACACTGACTCAAAAGTCAAGGT 109916
493 LeuAlaGlnHisIleHisAlaLeuThrLeuSerProSerAlaAlaHisTh 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109915 CTGCGTCAACACAGACACACTCTTAAGTTCCTCAGCAGCTCACAC 109866
509 rLeuAspLeuLeuHisAspHisProHisProGluGluAlaGluAla 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109865 TCTCGCTCTTCTTCACGATCATCCCATCTCCAGAGAGGACAGAGCAA 109816
526 rGSerAlaMetIleValHisMetProAspLeuMetAspIleAla..... 540
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109815 GATTCGCAATGACCATTCATGTCAGATCTTCATGATATACAGATCTA 109766
541 .....HisAlaArgSerArg 545
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109765 GATCAAGTCCCTCCCTTGATTCATGATATCAACAGATCTAGGTCAAG 109716
545 gSerProProTyrArgArgTyrArgSerArgSerArgSerProProGluP 562

```

```
|||||.....|
109715 GTCCCTCCCTATAGACATACCACATGCTCCAGATCTCTCCAGCAT 109666
562 heArgGlyGlnSerProThrLysArgAsnValProArgGluGluLysGlu 578
|||||.....|
109665 TTAAGGAACAGTCTCCACTAAATCAATGTCCTCAAGA..... 109625
579 ArgGluTyrPheAsnArgTyrArgGluValProProProTyrAspIleIy 595
109625 ..... 109625
595 sAlaTyrTyrGlyArgSerValAspPheArgPropheGluLysGlu 612
|||||.....|
109624 .....GAGAGAGAGC 109615
612 IgtYrArgGluTrrGluArgLysTyrArgGluTrrPyrGluLysTyrTyr 628
|||||.....|
109614 GCTACCAGCAATGGAGAGAGCACAGGGCGGTATGAGAT...AGG 109568
629 .LysGlyTyrAlaValGlyAlaGlnProArgProSerAlaAsnArgGlu 645
|||||.....|
109567 CAACGGCAGCAGTAGAAGCACTTCTAGACCTTAGGCCAAGAGAGAG 109518
645 sPheSerProGluArgLeuLeuProLeuAsnIleArgAsnSerProphe 661
|||||.....|
109517 TCTTTCTCCAGAGAGAACTTACTCTTAATATCATAAACCCACCTTC 109468
662 ThrArgGlyArgArgGluAspTyrAlaAlaGlyGlnSerHisArgAsnAr 678
|||||.....|
109467 CCAGAGGGCACGAGAGACTATGTCTGGCAAGTCGTAGAAATCA 109418
678 gAsnLeuGlyGlyAsnTyrProGluLysLeuSerThrArgAspSerHis 695
|||||.....|
109417 A.....GAGGCAACTACCCAGAAAGTTTCAACAGACACAGTCACA 109374
695 snAlaLysAspAsnPro.Lys.....SerLysGluLysGluSerGluAs 709
|||||.....|
109373 ATGCAAAAGATATATCCAAAGAGATTAACCAAAAGAGAAAGTGAGAA 109324
709 nValProGlyAspGlyLysGlyAsnLysHis 719
|||||.....|
109323 TCTTCAGGAGACAGCAAAAGAACACAT 109293
```

U.S. PATENT AND TRADEMARK OFFICE
PAGE BLANK (USPTO)

alignment_scores:
 Quality: 5635.50 Length: 1450
 Ratio: 4.403 Gaps: 18
 Percent Similarity: 88.276 Percent Identity: 77.379

alignment_block:
 US-09-811-045a-1 x AAI60201 ..

Align seg 1/1 to: AAI60201 from: 1 to: 4871

```

2 MetGLuValIysAspProAsnMetLysGLyAlaMetLeuThrAsnThrGI 18
|||||
163 ATGGAAGTGAAGAAGATCTAATATGAAGGTGCATGCTTACCAACACTGG 212
18 YLSTYrAlAlleProThrIleAspAla.GluAlaTYrAlAlleGLyLys 34
|||||
213 AAATATGCAATACCAACTATATGATGCAGAGAATATGCAATGGGAG 262
35 LysGLuLysProProPheLeuProGLuGLuProSerSerSerSerGLuGI 51
|||||
263 AAAGAAGAACTCCCTCTTCTACAGAGAGGCACTCTTCTCTAGAAAGA 312
51 uAspAspProIleProAlaGLuLeuLysLeuIleCysLysAspIleu 68
|||||
313 AGATGATCCATCCAGATGAATGTTGTCTCATCTGCAAGGATATTA 362
68 eThrAspAlaValAlleProCysCysGLyAsnSerSerCysAspGLu 84
|||||
363 TGACTGATGCTGTGTGATTCCTGCTGTGGAAACAGTACTGTGATGA 412
85 CysIleArgThrThrLeuGLuIleSerAspLysThrCysProThrCys 101
|||||
413 TGTATAGAAAGACGACTCTCGAATGCAGATGACGACACATCCGAGCTG 462
101 sHisGLnAsnAspValSerProAspAlaLeuIleAlaAsnLysPheLeu 118
|||||
463 TCATCAAAATGATGTTCTCTGATGCTTATATGCCAATAAATTTTATC 512
118 rGlnAlaValAsnAsnPheLysAsnGLuThrGLyTYrThrLysArgLeu 134
|||||
513 GACAGGCTGTAATTACTTCAAAAATGAAGTGGCTATACAAAAAGACTA 562
135 ArgLysGLuLeuProProPheLeuPheLeuValProProAlaArgProLe 151
|||||
563 CGAAAGAGGTACCTTCTCCACACACCCCAATACACCTCCGAGACACT 612
151 uSerGLnArgAsnLeuGLnProAlaArgSerArgSerProIleLeuArgLys 168
|||||
613 GATTCAGAGGAACCTACAACTGTGATGATCTCCGATATCAAGACAAC 662
168 LAspProValAlaPheArgTYrThrValSerProThrCysSerAspThr 184
|||||
663 AAGATCCTCTATGATTCACAGTACATCTCATCACT..... 700
185 LysThrAlaGLySerCysSerAspSerGLyThrLeuSerArgLeuProAl 201
|||||
701 .....CACCCAGC 708
201 aProSerIleSerSerLeuThrSerAsnGLnSerSerSerLeuAlaProProV 218
|||||
709 TCCGCTATATCTTCATTAACTTCTTAATCAGCTTCCTGGCCCTCTCG 758
218 aLserGLyAsnProSerSerAlaProAlaProValProAspIleThrAla 234
|||||
759 TGTCTGGAAATCCGCTTCTGTCTCAGCTCCGTGACTGATTAATACCTGA 808
235 ThrValSerIleSerValHisSerGLuLysSerAspGLyProPheArgAs 251
|||||
809 ACAGTATCCATATCATTCATTCAGAAAAATCAGATGAGACCTTTTGGGA 858
251 pSerAspAsnLysLeuLeuProAlaAlaLeuThrSerGLuHisSerL 268

```

```

859 TTCTGATTAATTAATATATTGCCAGCTGCAGCTTGTGCATCAGAGACTCA 908
268 YsGLyAlaSerSerIleAlaIleThrAlaLeuMetGLuGLyGLyVal 284
|||||
909 AGGGAACCTCTCATTTGCATTAACCGCTCTTATGGAAGGAAGGCTTA. 957
285 ProGLyThrSerProThrAsnSerIlePhe.ValGLyGLnSerLeuLeuH 301
|||||
958 CCAAGTCCCTGTTCTTGTGAACCCCATCTTGTGTGGACAGTATATTGC 1007
301 IsGLyGLnLeuIleProThr-ThrGLyProValArgIleAsnAlaAlaArg 317
|||||
1008 ATGGACAGTTGATCCCAACACTGTGTCAAGTAAATTAATACGTCTGT 1057
318 ProGLyGLyGLyArgProGLyTYrGLuHisSerAsnLysLeuGLyTYrLe 334
|||||
1058 CCAGGTGTGTGTGCACAGGCTGGGAACATTTCCAAACAACCTGGCTATCT 1107
334 uValSerProProGLnGLnIleArgArgGLyGLuArgSerCysTYrArg 351
|||||
1108 GGTTCCTCCACACACAAATTAAGAGAGGAGAGGCTGTCTACAGAA 1157
351 eIleAsnArgGLyArgHisSerGLuArgSerGLnArgThrGlnSer 367
|||||
1158 GTATAAACCGTGGCGACACACAGCAAGAGATCAGAGAGACTCAAGGC 1207
368 ProSerLeuProAlaThrProCysPheValProValProProProLe 384
|||||
1208 CCGTCACTACAGCACTCCAGCTTTGTACTGTCCACACACTCTCTT 1257
384 uTYrProProProProHisThrLeuProLeuProProGLyValProProP 401
|||||
1258 GTATCCGCTCTCCCATACACTTCTCTCCCTCCGAGGTGTTCTCTCTC 1307
401 rGlnLysSerProGLnPheProSerSerGLnPro.ProThrAlaGLyTY 417
|||||
1308 CACAGTTTCTCTCATGTTTCTCTCTGTGCGCAGCCAGCACCCGCTGGGA 1357
417 rSerValProProProGLyPheProProAlaProAlaAsnIleSerThra 434
|||||
1358 TAGTGCTCCTCTCCAGAGGTTTCTCCAGCTCCGCCAATTTATCAACAC 1407
434 LAcYsPheSerProGLyValProThrAlaHisSerAsnThrMetProThr 450
|||||
1408 CTGGGATTCATCAGAGGTGCAGACGCTCATCAATCCATCCCAACA 1457
451 ThrGlnAlaProLeuLeuSerArgGLuGLuPheTYrArgLysGLn.... 465
|||||
1458 ACACAGACACACCTTGTCCAGGGAAGAAATTATAGAGAGACGACAG 1507
466 .....AsnA 467
1508 ACTAAAGAGAGAGAAAAAGAAAGTCCAGCTAGATGATTATCAAAATG 1557
467 sPLysGLyArgGLu..... 471
1558 ATTTGCTTAAGGAATGATGGAATACAAAAGATTCAAAAGAGCGTATGG 1607
472 .....SerLysPheProTYrSerGLySerSerTYrSerAr 483
1608 CGGTCAATTTCCAGAGTCTTAATTCCTATAGTGTCTTCTGTAATTCAG 1657
483 gSerSerTYrThrAspSer.SerGLnGLyLeuAlaGLnHisIleHisAla 499
|||||
1658 AAGTTCATATACTTAATCAAGATCTGTGTCACACAGTTCACACT 1707
500 LeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHisAspPhe 516
|||||
1708 CTATTTCTGATCATTCAGCGCTCACATTTCTGTCTCTATTCACAGGCTCA 1757
516 sProHisProProGLuGLuAlaAlaArgSerAlaMetIleValHisM 533
|||||

```

1758 CCTCCATACCCAGAGAGGAGGCAAGCCGCAATTCACGTTACAG 1807
533 etProaspLeuMetAspIleAlaHisAlaArgSerArgSerProProtyr 549
1808 GTCATATCTCATGATATCAT. CGATCTAGGTCAGAGTCACCCCTTAC 1856
550 ArgArgTyrArgSerArgSerArgSerProProGluPheArgGlyIle 566
1857 AGACGCTATCATTCACGATCAGATCTCCTCAAGCCTTAAAGGACAGTC 1906
566 rProthrLysArgAsnValProArgGluGluLysGluArgGluTyrPhe 583
1907 TCCTAATAACGTAATAGTACTCAAGGGAACAGAACTGAATATATTAA 1956
583 sNArgTyrArgGluValProProProTyrAspIleLysAlaTyrTyrGly 599
1957 ATAGATACAGAGAGATTCACACCATATGACATGAAGCATATATAGG 2006
600 ArgSerValAspPheArgAspProPheGluLysGluArgTyrArgGluTyr 616
2007 AGAAGCTTGACTTATAGACCATTTGAAAAAGAGCTACCGAGAAATG 2056
616 pGluArgLysTyrArgGluTyrTyrGlyLysTyrTyrLysGlyTyrAla 633
2057 GGAGAGAAATATATAGAGCTGATGAAAAATATATTAAGGTTATGCTG 2106
633 aLGIyAlaGluProArgProSerAlaAsnArgLysAspPheSerProGlu 649
2107 CTGGAGCACAGGCTAGACCTCAGCAAAATAGAGAGAACTTCTCCAGAG 2156
650 ArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArgGlyLysArg 666
2157 AGATTTTTCACCTTACATCAGGAATTCCTCCTCACAAGAGGCCGACAG 2206
666 gGluAspTyrAlaAlaGlyLysSerHisArgAsnArgAsnLeuGlyAla 683
2207 AGAAGACTATGTGTGGGCAAGGTCATAGAACTGAAACATAGTAGCA 2256
683 sNtyrProGluLysLeuSerThrArgAspSerHisAsnAlaLysAspAsn 699
2257 ACTATCCAAAAAGCTTTCAGCAAGAGATGTCACATCAGAAAGGATTAAT 2306
700 ProLysSerLysGluLysGluSerGluAsnValProGlyAspGlyLysGly 716
2307 ACAAAAGTCAAAAGAGAGAGAGAGTGAAGCGCTCCAGAGATGTGAAGG 2356
716 yAsnLysHisLysLysHisArgLysArgArgAsnGluLysGlyLysGlu 733
2357 AATAAGCATTAAGAAACACAGAAAAAGAGA.....AAAGGGAGAG 2397
733 LysSerLysSerPheLeuAsnProGluLeuGluLysSerArgLysCys 749
2398 AAAGTGAGGGTTTCTGAACCCGAGCTTATAGAGACTTCTAGAAATCA 2447
750 ArgGlySerSerGlyLysAspGluThrLysThrAspThrLeuPheValle 766
2448 AGAGAACCTACAGAGGTGTGAAGAAATTAACAGACTCATGTGTGTCT 2497
766 uProSerArgAspAspAlaThrProValArgAspGluProMetAspAla 783
2498 CCCAACTAGAGATGATGCACACACTGTAGAGATGAACCAATGATGACAG 2547
783 LysSerLeuThrPheLysSerValSerAspLysAspLysArgGluLysAsp 799
2548 AATCAATCACTTTTAATCAGTCTGAAGAAAGACAGAGAGAAAGGAT 2597
800 LysProLysValLysSerAspLysThrLysArgLysSerAspGlySerAl 816
2598 AAACCAAAAGCAAGAGGTGATTAACCAACCAAGCAATATGATGATCTGC 2647
816 arThrAlaLysLysAspAsnValLeuLysProSerLysGlyProGluGlu 833
2648 TGCTCCAAAAAGAAATATGTATAAACTCTAAAGAGCCCAAGAGAA 2697

833 yValAlaSpGlyAspArgGluLysSerProArgSerGluPro. ProLeuTy 849
2698 AAGTATATGAGAACCTGAGAGATCTCTCGATCTGAACTCCCAATTAA 2747
849 sLysAlaLysGlu...GluAlaThrLysIleAspSerValLysProSerS 865
2748 AAAGCCCAAGAGAGACTCCGAAAGACTGTGACATTAATCAATCAAT 2797
865 eSerSerGluLysAspGluLysValThr. GlyThrProArg. LysAlaH 881
2798 CTTCCTCTCAGAAAGATGAAAAATCATCTGGAAACCCAGGAAAAAGCTC 2847
881 fSerSerSerAlaLysAsp. ThrArgArgGlnSerGlnProArgThr 897
2848 ACTTAAATCAGCAAAAGAACACCAACCAACCAACCAAGCAAGAGAA 2897
897 gArgSerLysArgThrValProLysThrSerSerGlnLysSerGlnPro 914
2898 AAAGTGAAGAGAGACTATTCCAAAAGATGCAAAATCAAGAAAGCTAAC 2947
914 aLArgThrArgArgProArgSerLeu...ArgLysIleAsnTyrLeuIle 929
2948 TAAAGAAAGAAAGGCCAGAGAGCTAATGAGAAAAACAAACCTTGATA 2997
930 AlaArgGluLysAsnGluArgGluLysArgLys. LysSerValAspLys 945
2998 ATAGAGAGAAAAAGAAAAAGAAAGAACTGAAAGAAAGAGCTAGATTA 3047
946 AspPheGluSerSerSerMetLysIleSerLysValGluGlyThrGlu 962
3048 GATTTGAGCTCTCTCAATGAAATCTGAAACTGAACTGAACTGAAAT 3097
962 eValLysProSerProLysArgLysMetGluLysAspValGluLysLeu 979
3098 AGTGAACCATCACCAGAAAGGCCAAATGCAACCTGATACGAAAAATGCG 3147
979 LuArgThrProGluLysAspLysIleAlaSerSerThrThrProAlaLys 995
3148 ATAGGACCCCTGAAAGAGACAAATTT...TCTTTAAGTGGCCAGCCAAA 3194
996 LysIleLysLeuAsnArgGluThrGlyLysLysIleGlyAsnAlaGlu 1012
3195 AAATCAACTCAACAGAGAACTGGGAAGAAATTTGGAAGTACAGAAAA 3244
1012 nAlaSerThrThrLysGluProSerGluLysLeuGluSerThrSerSer 1029
3245 TATATCAACACAAAAAGAACCTCTGAAAAATGAGTCAACATCTAGCA 3294
1029 ySILeLysGlnGluLysValLysGlyLysAlaLysArgLysValAlaGly 1045
3295 AAGTTAAACAGAAAAAGTCAAAAGGAAAGGTCAAGCAAAAAGTACTGGA 3344
1046 SerGluGlySerSerSerThrLeuValAspTyrThrSerThrSerSer 1062
3345 ACTGAAGGATCCAGCTCAACTCTGGTGATTTACACAGTACGAGCTCAAC 3394
1062 rGlyGlySerProValArgLysSerGluGluLysThrAspThrLysArg 1079
3395 TGGAGGAGCTCTGTGGGAAATCTGAAGAAAAACAGATACAAAGGAA 3444
1079 hrValIleLysThrMetGluGluTyrAsnAspAsnThrAlaProAla 1095
3445 CTGTGATTAACAGATGGAAGAAATATATATATACATATACGGCCAGCT 3494
1096 GluAspValIleIleMetIleGlnValProGlnSerLysTrpAspLys 1112
3495 GAAGATGTTATCATTTATGATTCAGGTTCCCAATCCAAATGGGATTAAGA 3544
1112 pAspPheGluSerGluGluLysAspValLysThrThrGlnProIleGln 1129
3545 TGACTTTGAATCTGAGAGAAAGATGTTAATCCACACACACTATATCA 3594

```

1129 eValAlGlySerProSerSerIleIleLysAsnValThrThrLysProSer 1145
|||||.....|
3595 GTCTAGGAAAACCTGCTAGTGTATATAAAATGTTAGTACAAAGCATCA 3644
1146 AlATrAlAlaLysThrGluLysGluSerGluGlnProGluLysLeuGln 1162
|||||.....|
3645 AATATAGTCAAGTATCTGAGAAAAGAAAGGAGCCATCCAGAAAATTTCA 3694
1162 nLysLeuProLysGluAlaSerHisGluLeuMetGlnHisGluLeuArg 1179
|||||.....|
3695 GAAATTCACCAAGACGTGACCATGAATCATACACATGAGGTTTAAA 3744
1179 eSerLysGlySerAlaSerSerGluLysGlyAlaGlyAlaLysAspArg 1195
|||||.....|
3745 GTTCAAAAAATCTGCTAGTGTAGTAAAAAGGAAACCAAGATCGAGAT 3794
1196 HisSerGlySerGluLysAspAsnProAspLysArgLysSerGlyAla 1212
|||||.....|
3795 TATTCAGTGTGGAAAAGAGAACCTGAAAAGGAGAAACAGCACTCA 3844
1212 nProAspLysGluSerThrValAspArgLeuSerGluGlnGlyHisPhe 1229
|||||.....|
3845 GCCAGAAAAGAGGTATTTGGACCGCTGAATGAAACAAAGAAATTTTA 3894
1229 yStrLeuSerGlnSerSerLysGluThrArgThrSerGluLysHisGlu 1245
|||||.....|
3895 AAAGTGTCTCATATCTTCCAAAGAGGCTAGAACGTCAATAAACATGAT 3944
1246 SerValArgGlySerSerAsnLysAspPheThrProGlyArgAspLys 1262
|||||.....|
3945 TCCCTGTGTCTTCTCAATTAAGACTTCCATTCAGAGACCAAAA 3994
1262 sValAspLysAspSerArgAspLysSerSerSerLysArgArgAspGlu 1279
|||||.....|
3995 AACTGACTATGACACAGAGAGTATTCAGTCCAAACGTAGAGATGAAA 4044
1279 rGgLyGluLeuAlaArgArgLysAspSerProProArgGlyLysGluSer 1295
|||||.....|
4045 AGAATGGAATTAACAAAGACGAAAGACCTCTCTCGAAATAAAGATTCT 4094
1296 LeuSerGlyGlnLysSerLysLeuArgGluGlnArgAspLeuProLys 1312
|||||.....|
4095 GCATCTGACAGAAAATTAACCAAGGAGAGAGATTTGGCTTAAAAA 4144
1312 sGly...AlaGluSerLysLysSerAsnSerSerProProArgAspLys 1328
|||||.....|
4145 AGGAACAGAGATTCACAAAAGAAAGTAAATTCAGTCCCAAGACAGAA 4194
1328 ySPronHisAspHisLysAlaProTyrGlnThrLysArgProGlyGlu 1344
|||||.....|
4195 AACCTCATGATCCAAAGCCACTTATGATTAACGGCCCAATGAGAG 4244
1345 ThrLysProValAspLysAsnSerGlyLysGluArgGluLysHisAla 1361
|||||.....|
4245 ACNAAATCTGAGTAAAAATCCTTGTAAGATGCTGAGAAACATGTTT 4294
1361 aGluAlaArgAsnGlyLysGluSerSerGlyAlaAsnCysHisValTyr 1377
|||||.....|
4295 AGAAGCAAGAAACAATTAAGAGTCAAGTGCATTAATACTTATTATAC 4344
1378 LeuThrArgGlnThrLeuProTyrPArgArgSerTrpLeuGlyArgTr 1394
|||||.....|
4345 TTAAACCCACAGAGACACAGGTTGAAAAGAGCAAAATTAATGCGCAATT 4394
1394 PArgArgAlaProSerSerArgAsnProSer 1404
|||||.....|
4395 GACAAGAGTACTGTCAMGCTTAACCCAGT 4425

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq - emb1/NA2001A.DAT:AA158415

seq_documentation_block:

ID AA158415 standard; cDNA; 4215 BP.
XX

```

AC AA158415;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 618.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
XX
PR 25-APR-2000; 2000US-0552317.
XX
PR 09-JUL-2000; 2000US-0598042.
XX
PR 19-JUL-2000; 2000US-0620312.
XX
PR 03-AUG-2000; 2000US-0653450.
XX
PR 14-SEP-2000; 2000US-0662191.
XX
PR 19-OCT-2000; 2000US-0693036.
XX
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dimaanic RT;
XX
DR WPI: 2001-442253/47.
XX
DR P-PDB: AAM39259.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX
PS such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 618; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening and
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 4215 BP; 1523 A; 901 C; 855 G; 934 T; 2 other;

```

alignment_scores:

Quality: 5381.00 Length: 1381
Ratio: 4.436 Gaps: 15
Percent Similarity: 87.835 Percent Identity: 77.190

alignment_block:

US-09-811-045A-1 x AA158415 ..

Align seg 1/1 to: AA158415 from: 1 to: 4215

```

1 MetMetGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsn 17
  |||
158 ATGATGAGAGTGAAGATCTTAATATGAAGAGGCAATGCTTACCAAC 207
  |||
17 rGlyLysTyrAlaIleProThrIleAspAlaGluAlaTyrAlaIleGly 34
  |||
208 TGGAAAATATGCAATACCAACTATAGATCAGAGCATATGCAATTGGGA 257
  |||
34 yLysGluLysProProPheLeuProGluGluProSerSerSerGlu 50
  |||
258 ACAAGAGAACCTCCCTCTTACAGAGAGCCATCTTCTCCAGAA 307
  |||
51 GluAspAspProIleProAlaGluLeuLeuCysLeuIleCysLysAsp 67
  |||
308 GAAGATGATCCTATCCAGATGATATGTTGTTCTCTATCTGCAAGATAT 357
  |||
67 eMetThrAspAlaValIleProCysCysGlyAsnSerSerCysAsp 84
  |||
358 TATGACTGATGCTGTGTGATTCCTGCTGTGGAAACAGTTACTGTGATG 407
  |||
84 lCysIleArgThrThrLeuLeuGluSerAspLysHisThrCysProThr 100
  |||
408 AATGTATAGAACAGCACTCTGGAATCAGATGAGCACACATGTCGACG 457
  |||
101 CysHisGluAsnAspValSerProAspAlaLeuIleAlaAsnLysPhe 117
  |||
458 TGTCACTCAAAATGATGTCTCTCTGATGCTTATATGCAATTAATTTT 507
  |||
117 uArgGlnAlaValAsnAsnPheLysAsnGluThrGlyTyrThrLysArg 134
  |||
508 AGCAGAGCGCTGAATTAATCTCAAAAATGAACGTGCTATACAAAAGAC 557
  |||
134 euArgLysGlnLeuProProPheLeuPheLeuValProProProArgPro 150
  |||
558 TACGAAACAGTTACTCTCTCCACACCCCAATACCACTCCGAGACCA 607
  |||
151 LeuSerGluArgAsnLeuGlnProArgSerArgSerProIleLeuArg 167
  |||
608 CTGATTCAGAGGACCTACACCTCTGATGAGATCTCCGATATCAAGACA 657
  |||
167 nGlnAspProValValPheArgTyrThrValSerProThrCysSerAsp 184
  |||
658 ACAAGATCCTCTTATGATTCAGATGACATCTTCATCAACT ..... 697
  |||
184 hLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPro 200
  |||
698 .....CACCA 703
  |||
201 AlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaPro 217
  |||
704 GCTCCGTATATCTTCAATTAATCTTAATCAAGTCTTCCCTGGCCCTCC 753
  |||
217 oValSerGlyAsnProSerSerAlaProAlaProValProAspIleThr 234
  |||
754 TGGTCTGGAATCCGTCCTTGCCTCCAGCTCTGATCCGATATAACG 803
  |||
234 lathValSerIleSerValHisSerGluLysSerAspGlyProPheArg 250
  |||
804 CAACAGTATCCATATCAGTTCATGAAAAAATCAGATGACCTTTTCG 853
  |||
251 AspSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerGluHis 267
  |||
854 GATTCGATATTAATTAATTCGACGCTGCACTTTCATCAGACGACCTC 903
  |||
267 rLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGluGluLysGly 284
  |||
904 AAAGGAACTCCTCAATTGCAATTAACGCTTATGAAAGAGAGGTT 952
  |||
284 aLProGlyThrSerProThrPasnSerIlePheValGlyGlnSerLeu 300
  |||
953 A.CCAGGTGCTGTTCTTGGAAACCCATCTTGTGTCGACAGTCATTA 1002

```

```

300 uHisGlyGlnLeuIleProThrThrGlyProValArgIleAsnAlaAla 317
  |||
1003 GCATGACAGTGTATGCCCAACTGCTCCAGTAAAGATTAATATCTGCTC 1052
  |||
317 rGProGlyGlyArgProGlyTTPGluHisSerAsnLysLeuGlyTyr 333
  |||
1053 GTCCAGGTGGTGTCCAGCAGGCTGGGAACATTCACAACACTTGGCTAT 1102
  |||
334 LeuValSerProProGlnGlnIleArgArgGlyGluArgSerCysTyr 350
  |||
1103 CTGGTTCTCCACCACACAAATTAAGAGGGAGAGAGCTGTACAG 1152
  |||
350 gSerIleAsnArgGlyArgHisHisSerGluArgSerGlnArgThrGln 367
  |||
1153 AATATTAACCGTGGGCGACACACAGCGAAGATACAGAGACCTCAAG 1202
  |||
367 eProSerLeuProAlaThrProCysPheValProValProProPro 383
  |||
1203 GCCCGTACTACAGACACTCCAGTCTTGTACCTGTCCACACCTCCT 1252
  |||
384 LeuTyrProProProProHisThrLeuProLeuProProGlyAlaPro 400
  |||
1253 TTGTATCCGCTCTCCCATACACTTCTCTCCCGGGTGTCTCTCC 1302
  |||
400 oProGlnPheSerProGlnPheProSerSerGlnProProThrAlaGly 417
  |||
1303 TCCACAGTTTCTCTCACTTCTCTCTGCGCAGCCAGCACCGCTGGGT 1352
  |||
417 ySerValProProProGlyPheProProAlaProAlaAsnIleSerThr 433
  |||
1353 ATAGTCTCCCTCCTCAGAGTTTCTCCAGCTCTCCCAATTTATCAACA 1402
  |||
434 AlaCysPheSerProGlyValProThrAlaHisSerAsnThrMetPro 450
  |||
1403 CCTGGGTATCATCAGAGAGTGCAGACGCTCATTAATAATCAATCCAC 1452
  |||
450 rThrGlnAlaProLeuLeuSerArgGluGluPheTyrArgGluGln .... 465
  |||
1453 AACCAAGCACCACTTTGTCCAGGGAAGATTCATAGAGACACGAC 1502
  |||
466 .....Asn 466
  |||
1503 GACTAAAGAGAGAAAAAAGTCCAAAGCTAGATGATTACAAAT 1552
  |||
467 AspLysGlyArgGlu ..... 471
  |||
1553 GATTTTGTAAAGAAATTGATGAATACAAAAGATTCAAAAGAGCGTAG 1602
  |||
472 .....SerLysPheProTyrSerGlySerSerTyrSer 483
  |||
1603 GCGCTATTTTCCAGGCTTAATCTCCCTTAGTGTCTCTGATATTCAA 1652
  |||
483 rGSerSerTyrThrAspSerSerGlnGlyLeuAlaGlnHisIleHisAl 499
  |||
1653 GAAGTTCATATCTTAATTAATCAAGATCTGGCAACACGTCGACGC 1702
  |||
499 aLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHisAsp 516
  |||
1703 TCTTAATCTCGATCATTCAGCCGCTCCTCAATTCCTGTTCTATTA 1752
  |||
516 lProHisProProGluGluAlaGluAlaArgSerAlaMetIleValHis 532
  |||
1753 ACCTCATACCCAGAGAGGCGAGGCAAGAGCCCAATTAACGTTTAC 1802
  |||
1803 GGTCTAGATCTCATGATATCATCGATCAAGGTCACACCCCTTA 1851
  |||
549 rArgArgTyrArgSerArgSerArgSerProProGluPheArgGlyGln 566
  |||
1852 CAGAGCTATCATTCACGATCAAGATCTCTCAAGGTTTAAAGGACAGT 1901
  |||
566 eProThrLysArgAsnValProArgGluGluLysGluArgGlyTyrPhe 582

```

[illegible]

| | | |
|------|--|------|
| 2793 | CCTCTCAGAAAGATGAAAAATTCACGTGACACCCCGAGAAAAGCTCACTCT | 2842 |
| 883 | LYSSerAlaLysASP..ThrArgArgInSerGlnProArgThrArgArgS | 899 |
| 2843 | AAATTCGCAAAAAGAACCCAGAAACAAACACACTCAAAAGAGAAAAGT | 2892 |
| 899 | eRLysrtrgThrValProLysThrSerSerGlnLysSerGlnProValArg | 915 |
| 2893 | GAAAGAGGACACTATTCCAAAGATGTCAAATTCAGAAAAGCTTAACACTAAG | 2942 |
| 916 | ThrArgArgProArgSerLeu..ArgLysIleAsnTryIleuLeuIleAla | 931 |
| 2943 | AAAGAAAGGCCAGAGAGCCTTAATGTGAGAAAACAAACCCCTTGATATAAG | 2992 |
| 931 | gGLyLSrASngLIaArgGluLysArgLys..LYSSerValAspLysAspH | 947 |
| 2993 | GGAGAAAAAGAAAAAGAAACCTGAAAGAAAAAGGCGTAGTAAGATTT | 3042 |
| 947 | eGLuSerSerSerMetLysIleSerLysValGluGlyThrGluIleValL | 964 |
| 3043 | TGAGTCTCTTCATGAATGAATCTCCAAACTGAAAGTGACTCAATTAATGA | 3092 |
| 964 | YsProSerProLysArgGlySMetGluLysAspValGluLysLeuGluArg | 980 |
| 3093 | AAACATCACCAAGGCCCAAAATGAAGACCTGATACAGAAAAATGATATGG | 3142 |
| 981 | ThrProGluLysAspLysIleAlaSerSerThrThrProAlaLysIle | 997 |
| 3143 | ACCCCTGMAAAGACAAATTT...TCTTTAAGTGGCGGCAAGCAAAAAAT | 3189 |
| 997 | eLYSLeuAsnArgGluThrGlyLysLysIleGlyAsnAlaGluAsnAlaS | 1014 |
| 3190 | CAAACTCAACAGAGAAACCTGGAGACAAATTTGGAAGTCAAGAAATATAT | 3239 |
| 1014 | eErhThrLysGluProSerGluLysLeuGluSerThrSerSerLysIle | 1030 |
| 3240 | CAACACAAAAAGAACCTCTGAAAAATTTGCACTCAACTCTAGCAAACT | 3289 |
| 1031 | LYSGluGluLysValLysGlyLysAlaLysArgLysValAlaGlySerG | 1047 |
| 3290 | AAACAGAAAAATCTCAAGAGAAAGTCAGACCAAAAGTGAAGCTGGAACTGA | 3339 |
| 1047 | uGLYSerSerSerThrLeuValAspTryThrSerThrSerThrGlyG | 1064 |
| 3340 | AGGATTCACAGCTCAACTGTGGGATTAACACAGTACGAGCTCACTGGAG | 3389 |
| 1064 | LYSerProValArgLysSerGluGluLysThrAspThrLysArgThrAl | 1080 |
| 3390 | GCAGTCTCTGTGCGGAAATCTGAGAGAAAAACAGATRCAAGAGGAATGTG | 3439 |
| 1081 | IleLysThrMetGluGluTryPAsnAsnAspAsnThrAlaProAlaGluAs | 1097 |
| 3440 | ATTAAACGATGAGAGATATTAATATGCAATACCGGCCGACGTGAAGA | 3489 |
| 1097 | pValIleIleMetIleGlnValProGlnSerLysTryAspLysAspAsp | 1114 |
| 3490 | TGTATCATTAATGATTCAGATTCCTCAATCCAAATGGCATTAAGATGACT | 3539 |
| 1114 | heGluSerGluGluGluAspValLysThrThrGlnProIleInserAla | 1130 |
| 3540 | TTTGAATCTGAGAGAGAAAGATTAAATCCACACAGCTATTAATCAAGTGA | 3588 |
| 1131 | GLYsPProSerSerIleIleLysAsnValLThrThrLysProSerAlaTh | 1147 |
| 3590 | GGAAACCTCTGATGTGTTATTAATAATTTAGTACAAAGCCATCAAAAT | 3639 |
| 1147 | rAlaLysTYrThrGluLysGluSerGluInProGluLysLeuGluLysL | 1164 |
| 3640 | AGTCAAGTATTCGAGAGAAAGTAGGACCAATCCGAGAAATTCAGAAAT | 3689 |
| 1164 | eupProLysGluAlaSerHisGluLeuMetGlnHisGluLeuArgSerSer | 1180 |
| 3690 | TCACCAAGGACGTGAGCCATGAATATATCAACATCATGAGTTAAAAAGTTCA | 3739 |

134 euArglysglnLeuProPheLeuPheLeuValProProProArgPro 150
|||||
778 TACGAAACAGTTACCTCCTCCACCACCCCAATACACCTCCGAGACA 827
131 LeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuArgI 167
|||
828 CTGATTTCAGAGAACCTACAACTCTGATGAGATCTCCGATATCAAGCA 877
167 nGlnAspProValIlePheArgTyThrValSerProThrCysSerAspT 184
|||||
878 ACAAGATCTCTTATGATTCAGAGTACATCTTATCAACT..... 917
184 hrLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPro 200
918CACCA 923
201 AlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaProP 217
|||||
924 GCTCGTATATCTTCATTAATTAATCTTAATCAGTCTCTGGCCCTCC 973
217 oValSerGlyAsnProSerSerAlaProAlaProAlaProAlaIleTh 234
|||||
974 TGTGCTGGAATCCGCTCTTCTGCTCCAGCTCTGTACCTGATTAACG 1023
234 IarhrValSerIleSerValHisSerGluLysSerAspGlyProPheArg 250
1024 CACAGATATTCATATCAGTTCATTCAGAAATATGATGAGACCTTTCG 1073
251 AspSerAspAsnLysLeuLeuProAlaAlaAlaLeuThrSerGlnHisSe 267
1074 GATTCTGTAATAATAATATGTCACCTCAGCTCTTGATGATCAGACACT 1123
267 rLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGluGluLysGly 284
|||||
1124 AAGGGAACCTCTCAATTGCAATTACGCTTATGAGAGAGAGGCTT 1172
284 alProGlyThrSerProThrAsnSerIlePheValGlyGlnSerLeu 300
|||||
1173 A.CCAGGTGCGTCTTGGAACCCCATCTTGTCGACAGTATTTAT 1222
300 uHisGlyLysLeuIleProThrThrGlyProValArgIleAsnAlaIa 317
|||||
1223 GCATGAGAGTTGATCCCAACATGTCAGTAAGAAATTAATCTGCTC 1272
317 rGProGlyGlyLysProGlyThrGluHisSerAsnLysLeuGlyThr 333
|||||
1273 GTCCAGGTGCTGACGAGCTGGAAACATTCACAAACTTGGCTAT 1322
334 LeuValSerProProGlnGlnIleArgArgGlyLysArgSerCysTy 350
1323 CTGGTTCTCCACCAACAATAATGAGAGGGGAGAGAGCTGCTACAG 1372
350 gSerIleAsnArgLysArgHisHisSerGluArgSerGlnArgThrGln 367
1373 AAGTATTAACCGTGGCGACACCAAGGAAAGATCACAGAGACTCAAG 1422
367 eProSerLeuProAlaThrProCysPheValProValProProPro 383
|||||
1423 GCCCGTCACTACAGCAACTCAGTCTTTGACCGTTCCACACCTCTCT 1472
384 LeuTyProProProHisThrLeuProLeuProProGlyValProP 400
1473 TTGTATCCGCTCCCTCCATACACTCTCTCCCTCGGGGTTCCTCC 1522
400 oProGlnPheSerProGlnPheProSerSerGlnProProThrAlaGly 417
|||||
1523 TTCACAGTTTCTCTCAGTTTCTCTCGGCGACCCACCCCTGGGT 1572
417 yrSerValProProProGlyPheProProAlaProAlaAsnIleSerThr 433
|||||
1573 ATAGTCTCTCTCCAGGTTTCTCTCAAGCTCTGCAATTATTAACA 1622

434 AlaCysPheSerProGlyValProThrAlaHisSerAsnThrMetPro 450
|||
1623 CTTGGTATATCATCAGAGTGCAGACAGCTCATTAATATACATCCAAC 1672
450 rThrGlnAlaProLeuLeuSerArgGluGluPheTyArgGluGln.... 465
1673 AACACAAACACACCTTGTCCAGGGAAGAAATTTATAGACAGCAGAC 1722
466Asn 466
1723 GACTAAAGAGAGAGAAAGAAAGATCCAACTGATGAGTTTCAAT 1772
467 AspLysGlyArgLys..... 471
1773 GATTTTGCTAAGAAATGATGAAATCAAAAGATTCAAAAGACCTGAG 1822
472SerLysPheProTySerGlySerSerTySer 483
1823 GCGCTCATTTTCCAGGTCTAAATCTCCTATATGTTGTTCTTCTGAT 1872
483 rGSerSerTyThrAspSerSerGlnGlyLeuAlaGlnHisIleHisAl 499
|||||
1873 GAAGTTCATATCTTATTTCTAATCAAGATCTGGGTCAACAGCTCAGC 1922
499 alLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHisAsp 516
1923 TCTTATCTCGATTCATTCAGCCGCTCAGATCTCTCTTATTCAGGTC 1972
516 isProHisProProGluGluAlaGluAlaArgSerAlaMetIleValHis 532
1973 ACCTTCATACCCACAGAGAGGAGGAGGAGCCGCAATTAACCTTCAC 2022
533 MetProAspLeuMetAspIleAlaHisAlaArgSerArgSerProTy 549
2023 GTCTA.GATTCATGATATATCAT.CGATCTAGCTAAGTCAACCCCTTA 2070
549 rArgArgTyArgSerArgSerArgSerProProGluPheArgGlyGln 566
2071 CAGACGTATCATTCACGATCAAGATCTCTCAACCGTTAGGGACAGT 2120
566 eProThrLysArgAsnValProArgGluGluLysGluArgGluTy 582
2121 CTCTTAATAAAGTAATGTACTCTCAAGGGAAGCAAGCGTAATTTT 2170
583 AsnArgTyArgGluValProProProTyArgAspIleLysAlaTy 599
2171 AATAGATACAGAGAAATTCACACCATATGACATGAAGATATTAAG 2220
599 yArgSerValAspPheArgAspProPheGluLysGluArgTyArgGlu 616
2221 GAGAAAGTGTGACTTTAGAGACCTTGTGAAGAAAGAACGTAACCGAAT 2270
616 rPGLuArgLysTyArgGluTyArgGluLysTyArgGluLysGlyTy 632
2271 GGGAGAGAAATATAGAGAGTGTATGAAATATTAATAAGTTATGCT 2320
633 ValGlyAlaGlnProArgProSerAlaAsnArgGluAspPheSerPro 649
2321 GCTGAGACACASCTAGACCTCAGCAATATAGAGAACTTTTCCAGA 2370
649 uArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArgGlyArg 666
2371 GAGATTTTGGCCACTTWTACATCMGAAATTCCTCTCAAGAGGCCCA 2420
666 rGGLuAspTyArgAlaGlyGlnSerHisArgAsnArgLysGly 682
2421 GAGAGACTATGTTGTGGCAAAAGTCAAGAGCGAAGATAGGTAGC 2470
683 AsnTyProGluLysLeuSerThrArgAspSerHisAsnAlaLysAsp 699
2471 AACTATCCAGAAAGCTTTCAGCAAGAGATGTCAATCAAGAGGATTA 2520
699 nProLysSerLysGluLysGluSerGluAsnValProGlyAspGlyLysG 716


```

1  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2521 TCAAAAGTCAAAAGAGAGAGAGTAAAAAGCTCCAGAGAGATGATRAAG 2570
716 LVSnlYshStlyLysLysHlsarGLysarGLysnglulYsglyglu 732
|||||
2571 GAAATTAAGCATAGAAGAACACAGAAAAAGAGA.....AAAGGGAG 2611
733 GluSerGluSerPheLeuasnProGluLeuGluThrSerArgLyscy 749
|||||
2612 GAAAGTAGAGGTTTCTGAACCCAGAGTATTAGAGACTTCTAGAAATC 2661
749 sarGLysSerSerGlyIleaspgluThrLysThrAspThrLeuPheVal 766
|||||
2662 AAGAGAACTACAGCTGTGAGAAATAATAACAGACTATTCTTTGTTTC 2711
766 euProSerArgaspaspalaThrProValargAspgluProMetaspala 782
|||||
2712 TCCCAAGTAGAGATGATGCCACACTGTAGATGAGATGACCAATGATGCA 2761
783 GluSerIleThrPheLysSerValSerAspglyAspglyargglulYsas 799
|||||
2762 GAATCATACACTTTAATCATAGTGTCTGAAGAAAGACAGAGAGAGAAAGGA 2811
799 pLysProLysValLysSerAspglyThrLysArgLysSerAspglySera 816
|||||
2812 TAAACCAAAAGCAAGGTGATTAACCAACGGAAGAGATGATGATCTG 2861
816 lathralalYsLysAspAsnValLeuLysProSerLysglProglnglu 832
|||||
2862 CTGTCTCCAAAAAGAAATATGTAAACCTGTCTAAAGAGAGCCCAAGAA 2911
833 LysValAspglyAspArgLysSerProArgSerGluProProleuY 849
|||||
2912 AAGGTAGATGAGACGTGAGAGATCTCCGATCTGAGACTGCAATTA 2961
849 slYsalalYsgluGluAlaThrLysLysSerValLysProSerSera 866
|||||
2962 AAAGGCCAAAGAGAGACTGCCAAGACTGACATCAATCATCATCT 3011
866 erSerGluLysAspgluLysValThrGlyThrProArgLysAlaHisSer 882
|||||
3012 CCTCTCAGAGATGAAAAAATCATCTGGAACCCCGAGAAAAAGCTCATCT 3061
883 LysSerAlaLysAsp..ThrArgArgInserGluProArgThrArgArg 899
|||||
3062 AAATTCAGCAAAAGACACCAAGAACAAACAGTCAAAGAGAGAAAGT 3111
899 erLysArgThrValProLysThrSerSerGluLysSerGluProValArg 915
|||||
3112 GAGGAAGGACTATTCCAAAGATGTCAATCAGAAAAAGCTAACAACTAAG 3161
916 ThrArgArgProArgSerLeu...ArgLysIleAsnThrLeuIleAla 931
|||||
3162 AAGAAAAAGCCCAAGAACCTTAATGAGAAAAACAAACCACTTGATTAAT 3211
931 ggLulYsasngluArgGluLysArgLys..LysSerValaspglyasph 947
|||||
3212 GGGAGAAAAAGAAAAAGAACTGAGAAAAAGCGCTAGATAAGATTT 3261
947 egluSerSerSerMetLysIleSerLysValGluGlyThrGluIleVal 964
|||||
3262 TGGAGTCTTCTCAATGAAAAATCTCGAAACTAGAGAGACTGAATATCTGA 3311
964 yspProSerProLysArgLysMetGluGlyAspValGluLysLeuGluArg 980
|||||
3312 AACCATTCACCAAGCCCAAAATGGAACCTGATCTGAAAAAATGGRTAG 3361
981 ThrProGluLysAspglyIleAla..SerSerThrThrProAlaLysLys 997
|||||
3362 ACCCTGAGAAAAAGCAAAATTTCTTTAAGTGGCCACC.....AAAAAA 3405
997 leLysLeuasnArgGluThrGlyLysLysIleGlyAsnAlaGluAsnAla 1013
|||||

```

```

3406 TCAAACTCAACAGAGAACTGGGAGAGAAATTCGAGAGTACAGAAATATA 3455
1014 SerThrThylYsgluProSerGluLysLeuGluSerThrSerSerLys11 1030
|||||
3456 TCAAACACAAAGAAAGAACCTCTGAAAAATTGAGTCAACATCTGCAAAAGT 3505
1030 eLysGluGluLys 1034
|||||
3506 TAAAAAAGAAAAA 3518
seq_name: /SID1/9cdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAS28668
seq_documentation_block:
ID AAS28668 standard; DNA; 17252 BP.
XX
AC AAS28668;
XX
DT 07-NOV-2001 (first entry)
XX
DE Genomic sequence #508 encoding for novel human respiratory antigen.
XX
KW Human; respiratory antigen; respiratory disorder; throat disorder;
KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
KW anti allergic; anti asthmatic; anti inflammatory; olfactory;
KW respiratory active; ds.
XX
OS Homo sapiens.
XX
WO200155448-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01333.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0188874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226886.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.

```


119 1n..... 119
11 AGGTACTGCTGATCATTTTATGAAAAAATCTTTTAACTGATTGA 6151
119 119
6152 ACTGACTTCAGTAGAATAGCATATTTCTGCAATTAATATGCTTG 6201
120AlaValAsnAsnPhelysAsnGluThrIglyThrLysA 133
6202 GATATCTGATGGGTGAATTAATTAAGTCAAAATGAACTGGCTATACAAA 6251
133 rGleuArIglysGlnLeuProPheLeuPheLeuValProPropArg 149
6252 GACTAGAAAACAGTTACTCTCCACACACCCCAATACCACCTCCGAGA 6301
150 ProLeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuAr 166
6302 CCATGATTCAGAGAACTACCAACCTCGATGAGATCCCGATATCAAG 6351
166 gGlnGlnAProValAlPheArgIyThrValSerProthCysSerA 183
6352 ACACACAGATCCTTATGATTCAGTACATCTCATCAACT..... 6394
183 sPThrIyThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeu 199
6395CAC 6397
200 ProAlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaPr 216
6398 CCAGTCCGCTCATATCTTCATTAACCTCTTAATCAGTCTCTTGCCCC 6447
216 oProValSerGlyAsnProSerSerAlaProAlaProValProAspIleT 233
6448 TCCTGCTGTGGAAATCCGCTTCTGCTCCAGCTCCTGATCAGATATA 6497
233 hrAlaThrValSerIleSerValIleSerGluIlySerAspGlyPro... 248
6498 CTGCACACATATCCATATCAGTTCAATTCAGAAAAATCAGATGACCTTTT 6547
248 248
6548 CGGTAGCCTGTGTGTTTTCACGTTAGAACAATAGAGTCATGAT 6597
248 248
6598 GTAGTGTGTTGTGTGTTATCATCTTTTAACAGCTTTCTTAGTGACCA 6647
248 248
6648 CTGTGCTCAGTAATGGGATGACTTTGAAGAAGTACTAGTCTTAATG 6697
248 248
6698 GGGGAGAGGAAAAACAACGTATGATTTGCAATGAGGATTAAAGAT 6747
248 248
6748 CCTGAGGTCAAAATACTACACAGGTGCAACACGCCCTATTAAAGCTTG 6797
248 248
6798 AATGTTCAATATCCATAACAGCTGGAACCTTAACAGAGGTGGAGAAAT 6847
248 248
6848 CACTTGATCATATGATAGTTTGGATAGAGTAGTTTGAAGTAAC 6897
248 248
6898 TAAGAACAGTTTAGATAGTAGAGAGGGGTAAAAACACAGATAAGT 6947

248 248
6948 AGAGAGTCAGGAATATGATGATTGTTTGAGCAATAATTAATATAGAGAG 6997
248 248
6998 AATAGTAAAAACTCACAGCTTACCACTTCAGAAATGAGACTTTGTGCAT 7047
248 248
7048 GATGTTAATGTCCTAATTTCTTTAAAAAAGGTGTAAAG 7097
248 248
7098 AGGCCATTTTCAGACACACTGAAGCAAAATCTTAGATGAGACCTTAC 7147
248 248
7148 ATTTACATTTAAGAAATTCACATGATCTGTGGAATTTATTTGGTGANA 7197
248 248
7198 GTGGAAATAGGCCAGAGAGTTCTATGGCTGACCCAGATTATCTTTCA 7247
248 248
7248 GTTTAGGTAGACATAGAGCTAAACTTAACCTTAATCCTAGGGGAT 7297
248 248
7298 TTTTGCTTCATAGGCCAGACTTTGATTTTAAGTTTACTATCTTGCAA 7347
248 248
7348 CTGATTTCTGTCAATACAGTAGATCTCTTAAGAGGTCCTTCATTGA 7397
248 248
7398 CTTTCCACTTACTACATATAGAGAACTCTCAAGTTTCAGGCTTCTCT 7447
248 248
7448 CACTTCTCAGAAATTCATCATCTTTGACTTAACATCTTAAGTGC 7497
248 248
7498 AGGTGATGGTAAACAGCAAACTTTTACTCTTAATGTCATTTTC 7547
249Phe..ArgAspSerAsp 254
7548 TGAATAATTTTAATGATTAATTAATTAATTTTAAAGGATTTCTGATA 7597
254 subLysLeuLeuProAlaIleAlaLeuThrSerGluIleSerLysGlyAla 270
7598 ATAAATATTTGCCAGCTGCAGCTTTGCATCAGACACCAAGGGAGACC 7647
271 SerSerIleAlaIleThrAlaLeuMetGlu..... 281
7648 TCCCTCAATTCGAATTAACCGCTTATGGAAGAGAAATTTTACTGG 7697
281 281
7698 TGCCTTAATTTGGGATTTTAAACAGAGGTTTGTCTTTCGGCTTT 7747
281 281
7748 GTTCTACCTTTTCTTTTCTTTTCTTTTAACTATTTCTGTAATTTGTC 7797
282LysGlyValProGlyThrSerProThrAsnSerIlePhe.V 295
7798 GTTGTCTTAAGGTTTACCGAGGTGCTGTCTTGGAAACCCCATCTTTCG 7847
295 aLysGlnSerLeuLeuHisGlyGlnLeuIleProThr..... 308

7848 TTTGACAGCATATTATTCATGAGACAGTTGATCCCACTGGTGGATTA 7897
308 308
7898 GATCAGCTTGGTTAGACCTTTTCCCTTTAAAAAATATTAGCTGA 7947
309G 309
7948 TTTAATGTCAATTTTGTATTACTAGCTGCTTTATTTTTGTCTAG 7997
309 lypOvalaIleasuaIalaIarProGIylGIyIarGProGIyTTP 325
|||||.....
7998 GTCAGTAAGAAATAATACGTCGTCAGGTGGTGGTGCACGAGCTGG 8047
326 GI..... 326
8048 GAACAGGAGTACATGTTTACAAATATCTCAGATGATTCCTGCCAAACT 8097
326 326
8098 AGATAATGAAGGTCCAACTAGGCAAACTAGATAATGAAGGTCCAAAC 8147
326 326
8148 TAGGACAGGGGTGCTGCTCTTATTGTAAACTTCATGTGCCAGATTAA 8197
326 326
8198 CATTTTTATGTAATAATGTCAAAATTCAGCTTATAGAGGCTTTT 8247
326 326
8248 AAATGACCAATCTTTCTAAATACAAAATTTTAAAGTTTCTTCTAC 8297
326 326
8298 TTTCAATCCACAANAATGGGAGTCTTGACAGAGCTTCAGTTTAATGAATA 8347
326 326
8348 GATAAGATTACTGAAAAGTTATTAAATATGTGGCAGTTATTTCC 8397
326 326
8398 TAGAGGTTTAACTTTTCTAACCTTAATTGGATGATCAAAAGTTGATT 8447
326 326
8448 TGAGTACATAAATACATPATAGATTAACTCTCCAAACAAAATTTT 8497
326 326
8498 TGCTGAATCGAGTTGTTCTGCCACCTAAATACAAAGAACTTGTAT 8547
326 326
8548 TTTCTGTGGGATATTTTATGCATTAATAATCAGCAGACTTACTTAAT 8597
326 326
8598 GTATTTTATGTAATAATATTCTGTTTATTAATCCGATGAAAAAT 8647
326 326
8648 GATATATTCTTAAAAAGTATGAGTTACTAGCAGAGATGATCAGTAGAAGA 8697
326 326
8698 ACATAGATTTAGAAACAGTAGACTTGGTTTGAATCCCAATCTCTTCA 8747
326 326

8748 TTTCTAGCTGCATGACAGTGTCTGTGTATTACTCCCTAACATCCAT 8797
326 326
8798 TTCCTGTTTGATTTGGAATGACACTACTACCTTAACATGTAAATGA 8847
326 326
8848 CTTCAATTGTGCCGACGACAAAAGGCTCTTTCTTGTATATAACAAAT 8897
326 326
8898 AAGTTCACAGATAATTTATAGAGTGTCTAATTAATAATATAATACATA 8947
326 326
8948 AAGTAATCATTTTGTAGAGTCTAGTTATTGTATGATACAGTTAATGGTG 8997
326 326
8998 GGGGAAAGTAGTTTATAGTAATAATATATTACTACGCTGTATCCG 9047
326 326
9048 AAACAATTTAAGACATGGGAGTAGSAAAAAGTTTCAGAGTAGTTAAA 9097
326 326
9098 ATTGTGCTTAAACAGCTATGATGAGAAAGTCATGTTATTAACCATGG 9147
327His.SerAsnIysLeu 331
9148 AGAAATCTTGAGCCTATTTTCTTCTCCACACAGATTCACAAACTT 9197
332 GIYTyIleuValSerProPogInIleargArgIylubArgSerCy 348
|||||.....
9198 GCGTATCTGTTCTCCACACAACAATTAAGAGGAGAGAGACTG 9247
348 sTyT..... 349
9248 CTACAGGTAGCATGTAAAAATCTGAAAATATATCATCTTATTTT 9297
349 349
9298 TATATATATCTCACCAAGAATAATTAAGCACAGCAGTGGCAGAAAGAG 9347
349 349
9348 GGGGTCAATTGTTGTATTGTGCTTATCCCTTAAGTCTGGGTAGTTG 9397
349 349
9398 TCCTTAAAGAGGCTAGAAAAGAAATTAATTAATATGATCTTAAATGA 9447
349 349
9448 TGTTTTTAACACATTTTAAATATTCCTCCAGTGAATTACTTCTCCCT 9497
349 349
9498 CTCAGCAACAGCAGAGATTTGGCTGTGATTTACCTCTTGGAAGTAGAA 9547
349 349
9548 GAAAGTCAGGGGTGACACAATATCTAAGATACAGTTTAATTAAGATTA 9597
349 349
9598 AAGTTAACAGCTCAGTCCCACTGTGAACAGAAATTCACATATTCT 9647
349 349
9648 CTCTAGACCCCACTGTCCCAATCATCTCCACGCTCTTTCCAAAAG 9697

349 349
9698 AAGACAGGAGATGTTAGAGCTCCTGCCCTAGTGGGGGTACAAAC 9747
349 349
9748 AATATTAACTTAAGAAGTAAAGATCTACTGTCTAGTCAAGGGTCCCA 9797
349 349
9798 ACCCTGGGCCACAGACAGATATCATCTTGGCTGTAGAACTGGGCA 9847
349 349
9848 ACACAGCAGAGGTGAACAGCAGCTGAGGAGACATTACCACTGAGCTCC 9897
349 349
9898 ACCTCTTGTAGATCAGCAGCTGGCATTTAGATTCTCATAGAGACATGAAC 9947
349 349
9948 CCTGTGTGAAGCTGGGTGCAAGAGATCTGGGTGCACGTTCTTAGGA 9997
349 349
9998 GAATTAATGCTGATCATCTAAGGTGGAAGATTTTACCCCCCTGCC 10047
349 349
10048 CCGTCATGAAAAAATGTCTTCCAAAAACAGCTCCCTGTGCCAAAC 10097
349 349
10098 GGTGAGAACTCTAGCTAGTCAATATCACACCAGATAAGAGGCC 10147
349 349
10148 TCCAAAACTTCAACAGAGTTTGGCCAGCATATGCTCATGCTGG 10197
349 349
10198 TAATTTAGAGCTTGGAGGCTGATGGGGAGGATGCTTGAGGCTGA 10247
349 349
10248 AGTTCAAGATTAGCTGGGCAACATAGCAAGACTGTCTACCAAAAA 10297
349 349
10298 AATAAAAATACAGACATGGTGGCATGCACCTGTGCTCCAGCTACCTG 10347
349 349
10348 GAGGCTGAGCAGAGATCATCTTGAGCTTAGAAGTCAAGGCTGCAATG 10397
349 349
10398 AGCTAATACCATCATCTGCACTTAGCCTAGGGAGCAGAGTGAACCT 10447
349 349
10448 GTCTCCAAAAAAGAAATGGGATATCTTCTAGTGTCTCATTTCTAA 10497
349 349
10498 ACTTATGAAGTTTACTATGTCTATTACTGATACATTTTAAAGAAGT 10547
349 349
10548 AATAGTTATTCATGGCTACGTGAAGAGATCTTAAATCTAAACCCCTAC 10597

349 349
10598 TGTGCTGTGTGTACCAATTAATTAATTTAAAGCACCTCTAGTTAA 10647
349 349
10648 CAGAAGTAACAGATCTCTGTTGGTGAGCACAATTAAGAACTTAGCCAA 10697
349 349
10698 CACTGTATTTTACTTAAGCTTTGAGAGAAAGCGTATATTTTCATT 10747
349 349
10748 GTTAATATAGAAGTTAAACTAGAAATTATCTGCTTTGTATGATCTT 10797
349 349
10798 ACCCACTTCCTCCCAAAAAAGAGGTACAGAGTTGGTTTCTTCATT 10847
349 349
10848 TTAAACTTTTGGGCCAGGTGTGTGCTCTCATCTTAACCCAGCACT 10897
349 349
10898 TTGTGAGATCACTGAGGTCAAGAGTTTGAGACACAGCTGCCAATG 10947
349 349
10948 GTGAACCTGTCTCTCTAATAAATACAAAAATTAAGTGGCATGGTGC 10997
349 349
10998 ACAAGTCTAGTCCAGCTACTTGGGAAGTGAGGAGAGAAATTCCT 11047
349 349
11048 TGAACCCAGAGGTGAGGTTCAGTGAAGCAAGATTACACCCTGACT 11097
349 349
11098 CCAGCTGAGCAGAGAGTGAGACTGTCTCAAAAAATAAATAAAT 11147
349 349
11148 GAATAAGTAAGCTTTGGGAGAGCTAGGTCAATAGCAGATTAACTGTC 11197
349 349
11198 AGTGCTTCAAAAAGCAGTTTAGAAGTAACAGGTAGTAATTAGTGACT 11247
349 349
11248 GATAATCACTGTAAGCTTCATAAAGAGAACTTTAAATACATATTTCC 11297
349 349
11298 GTGTAGAGAAAGTCTATTAACCTTGATCATCTCAGTCGAGATGGCT 11347
349 349
11348 TAAAAATTATGTTGTAGAGTTATTTTACAAATAAATGTACTAATGTG 11397
349 349
11398 CTACACTAGTGAATAGTTGGTTCTATTCACAGTTTAAAGTTGAAAA 11447
349 349
11448 TAATTAATGATGAGAGATTAAGCTTACTTGTCTCAGATGACTTTAGTT 11497
350 ArgserIleasnArg1A 356

```
11498 GAAGAGTAAATATCTTGGAAATTTATTTTCCAGAGTAAACCGTGGC 11547
356 rghIshISerGIuArSergInArGTThGIInSerProSerIeuProAla 372
11548 GACACACAGCGAAGATCAGAGAGACTCAAGGCCCTCCTACTACACAGA 11597
373 ThrProCysPheValProValProProProIeuTyProProProPr 389
11598 ACTCCAGCTTTGTACTGTTCACACACCTCCTGTGTATCCGCCCTCC 11647
389 ohISThIeuProIeuProProGIyValProProProGIInPheSerProG 406
11648 CCATACACTCTCTCCCTCCGGGTTCCTCCACAGTTTCTCTCC 11697
406 InPheProSerGIInProProThAlaGIyTySerValProProPro 422
11698 AGTTTCCCTCCAGCCAGCCACCCCGCTGGGTATAGTCCCTCTCCA 11747
423 .GIyPheProAlaProAlaAsnIleSerThAlaCysPheSerProG 439
11748 GGGGTTCCTCCAGCTCCTGCCAATTTATCAACACCTGG .GTATCATCAG 11796
439 IyValProThAlaHisSerAsnThMetProThThGIInAlaProIeu 455
11797 GAGTGCAGACAGCTCATTCAAATACCATCCACAAACAGACACACCT 11846
456 LeuSerArg .GIuGIuPheTyArgGIuIn. 465
11847 TTGTCCAGGGAGAAATCTCTATAGAGACAGCAGACTAAAGAGAGT 11896
465 11897 AATGATCTAAATTTGAATATATATACACTTTTTCATTTCTGTATAA 11946
465 11947 ACATTAATAGGTGTCTGAAAAAGTTTCCAGTGTTCAGTGTAGTATGA 11996
465 11997 GAAATTCACCTAAATTTAAACCAAGCCATACAGCAACTTTCAGGGA 12046
12047 TTCCAGTATTATAGTTTAAAGTGGAAACTAATGCAATTCATTCAAT 12096
465 12097 TCATTCATGATCTATATACATTTGTCTGTAGTTTCTTAATGGTTTT 12146
465 12147 TGTTAACTTTCACTCTTTAACTTATTTACTAGGAAAAAGAAAGTCC 12196
465 12197 AAGCTAGATGAGTTACAAATGATTTTGTAAAGAAATGATGAAATACA 12246
466 ..AsnAspIySgIy. 469
12247 AAAGATTCMAAAGAGAGCGCTAGCGCTCATTTTCCAGTGTAGTGTTCGA 12296
469 12297 AGCCTTCAACAAGAAATACAAACGGAGCTTTGAATATCCAGGATTAAC 12346
469 12347 TATGATATTAAGAACAATGCATTTAGGAATTTTCTGATTCGGTCTAGG 12396
469 12397 TGTATAGAAAGTGAAGGTACCTTATTTCCAAATATCTGTTCAATGAAT 12446
12447 AATGCTCAGTGAATTTTATGTGTAAGTAACAAGTTGTGGGATTA 12496
469 12497 AATTGCCCTGTCCCTTAAGTTATTCATAGTATCAATAATATGGAAAT 12546
469 12547 GTTTTGGAGAAAGAAATATGGGAAATAGCCAAAAACAGTTCCACAGA 12596
469 12597 CCCTTTACATAAGGCTACAGCTTCTGTGTGAACATTAACCTGCCATA 12646
12647 TTATGAAGTAGTAGTAGAATACAGATCCGTTATGTATGATAAAAAAGTGA 12696
469 12697 TTATTAATAATTCCTCGAAGTTGTTTAAAGACTTATTTTATGACCTT 12746
469 12747 ACCTGAAATGTGTTCTTTTAAACTAGGCTGTGGCCAGTATAGA 12796
12797 CAAAAACAGACTCCTACTTCATATTCAGATAGAGAAAGTAAATTA 12846
469 12847 TTAATGACATTTGTTAATGTTAGTTAGCTCATATATACACAGGTACA 12896
469 12897 TTAATGACACTTACCTTAATGTTAAGACTTACACTTTGTTTGTAAACA 12946
469 12947 GTGATGCTCCGATTTGTACTTATTCATTCACAGTAGAGTCACTGT 12996
470 12997 GCTCTTAAATCATCTAAGATGAATTTGATTAAGATAGATGAACCG 13046
470 13047 TGAGGAAGAATGCTAGAATCTGAGAGTGTGTTTATTTGTATGATAT 13096
472 13097 TTCTCACTATCAACTATTTGTTTGTGTTTGTAGGCTAATAATCTCC 13146
476 13147 TATAGTGTCTCTGATTTCAAGAGTTCAATATCTTATTTAATCAAG 13196
492 13197 ATCTGTTCAACAGCTTACCGCTCTTATCTGATCATTCAGCGGCTCAC 13246
509 13247 AATCTCGTTCTTATTCACGCTCACCTCCATACCCAGAGAGGAGAGG 13296
525 13297 AAGAGCCGCAATTAACCTTACAGGCTAGATCTCATGATATTCAT .CGAT 13345
```



```

15134 AGAATTAAGTTTGTGTTTTTTTAAATGTTAAAGACATTTGTGTTTAAAG 15183
1058 ..... 1058
15184 TTACGATTTTATATTTATTCAGCTGTTTATATTAATTAATGTAATTCCTCA 15233
1059 .....ThSerSerThGly 1064
15234 TTAATTAATAATTTTGTATTCCTTTTACAGAGCTCAACTGAGAGC 15283
1065 SerProValArgLysSerGluGluLysThrAspThrLysArgThrValI 1081
15284 AGTCCTGCGGGAATCTGAAGAAAAACAGATCAAGACGAACTGTGAT 15333
1081 eLysThrMetGluLysTrpAsnAspAspThrAlaProIaGluAspV 1098
15334 TAAACGATGGAAGATTAATTAATGACAAATACCGCCACCTGAAAGATG 15383
1098 alIleIleMetIleGlnValProGlnSerLysTrpAspLysAspPhe 1114
15384 TTATCATTTATATTCAGCTTCCTCAATCCAAATGSGATTAAGATGACTTT 15433
1115 GluSerGluGluGluAspValLysThrThrGlnProIleGlnSerValG 1131
15434 GAATCTGAAGGAAGATGTTAAATCCACACAGCCTATATCAAGTGTAGG 15483
1131 yLysProSerSerIleIleLysAsnValThrThrLysProSerAlaThr 1148
15484 AAAACCTGCTGCTGTTATTAATAAATGTTAGTACAAAGCCATCAATTAAT 15533
1148 IaLysTrpThrGluLysGluSerGluGlnProGluLysLeuGlnLysLeu 1164
15534 TCAAGTATCCCGAGAAAGATGAGCCATCCGAGAAATTCAGAAATTC 15583
1165 ProLysGluAlaSerHisGluLeuMetGlnHISGluLeuArgSerLys 1181
15584 ACCAAGGAGCGAGCCATGAATCATACACATGAGTTAAAGTTCAAA 15633
1181 sGlySerAlaSerSerGluLysGlyArgAlaLysAspArgGluHisSerG 1198
15634 AAACCTGCTCATCTAGTGAAGAAAGGAAACCAAGATGAGATTAATTCAG 15683
1198 LysSerGluLysAspAsnProAspLysArgLysSerGlyAlaGlnProAsp 1214
15684 TGTGGAAGAAAGAGAACCTGAAAGAGAAAGAACAGCACACGACGACAG 15733
1215 LysGluSerThrValAspArgLeuSerGluGlnGlyHisPheLysThrLe 1231
15734 AAAAGAGATTAATTTGACCGCTGATGAAACAGAAATTTTAAAGTCT 15783
1231 uSerGlnSerSerLysGluThrArgThrSerGluLysHisGluSerValA 1248
15784 GTCTCAATCTTCCAAAGAGCTAGACGTCAGATTAACATGATTCACCTC 15833
1248 rGlyLysSerSerAsnLysAspPheThrProGlyArgAspLysLysValAsp 1264
15834 GTGCTCTCTCAATTAAGACTTCACTCCCAATAGACAAAAAACTGAC 15883
1265 TyrAspSerAlaGAspTrpSerSerLysArgArgAspGluArgGlyG 1281
15884 TATACACACCAAGAGATTAATCAAGTTCCAAACGTAGAGTGAAGAAAGA 15933
1281 uLeuAlaArgArgLysAspSerProProArgGlyLysGluSerLeuSerG 1298
15934 ATTAACAAGACGAAGAACTCTCTCTCGGAATTAAGATTCGTGATCTG 15983
1298 LysGlnLysSerLysLeuArgGluGluArgAspLeuProLysLysGly... 1313
15984 GACAGAAAAATTAACCAAGGAGAGAGATTTGCCATAAAAAAGGAGACA 16033
1314 AlaGluSerLysLysSerAsnSerSerProProArgAspLysLysProH 1330
:::.....

```

```

16034 GGAGATTCCAAAAAAGTAATTCATGCTCCCTCAAGACACAGAAAACTCA 16083
1330 sAspHisLysAlaProTyrGluThrLysArgProCysGluGluThrLysP 1347
16084 TGATCCAAAGCCACTTATGATACAAAGCCCAATTAAGAGACAAAT 16133
1347 rovalAspLysAsnSerGlyLysGluArgGluLysHisAlaIaGluAla 1363
16134 CTGTAGATTAATAATTCCTGTAAAGATCGTGAAGAACATGATTAAGACA 16183
1364 ArgAsnGlyLysGluSerSerGly AlaAsnCysHisValTyrLeuThr 1380
16184 AGGAACAATAAAGAGTCAAGTGCAATTAACCTTTATATCTTAACCC 16233
1380 rGlnThrLeuProTyrParGArgSerTrpLeuGlyArgTrpArgArg 1396
16234 ACCAGAGACACAGTTGAAAAAGACAAATTAATCTGCGCAATTTGACANA 16283
1397 AlaProSerSerArgAsnProSer 1404
16284 GTACTGTCACGCTAAACCCAGT 16307
seq_name: /STD1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:AA067601
seq_documentation_block:
ID AA067601 standard; cDNA; 1800 BP.
XX
AC AA067601.
XX
DT 20-NOV-1994 (first entry)
XX
DE Retinoblastoma protein Ap4.
XX
KW Retinoblastoma-associated polypeptide; RAP; Ap12; Ap4;
KW transcription factor; E2F; cell cycle; ss.
XX
OS Homo sapiens.
XX
PN MO9412521-A.
XX
PD 09-JUN-1994.
XX
PF 19-NOV-1993; 93WO-US11310.
XX
PR 20-NOV-1992; 92US-0979156.
XX
PA (TEXA ) UNIV TEXAS SYSTEM.
XX
PI Lee W, Shan B;
XX
DR WPI; 1994-200185/24.
XX
PT Nucleic acid encoding retinoblastoma-associated polypeptide(s) -
PT used for producing prods. for use in studying cell cycling and
PT dysregulated cell growth.
XX
PS
XX
XX Disclosure; Page 37-38; 77pp; English.
XX
CC CDNA encoding a retinoblastoma (RB) associated protein (RAP), Ap12,
CC was cloned by direct screening of cDNA expression libraries using
CC purified RB protein as probe. Ap12 showed transcription factor E2F
CC and RB binding activities. The 5' and 3' cDNA sequences were
CC determined for Ap2 (AA067395, AA067596), Ap8 (AA067597, AA067598) and
CC Ap15 (AA067599, AA067600), and full-length encoding sequences for Ap4
CC (AA067601) and Ap10 (AA067601), which also encoded RB-associated
CC proteins.
XX
SQ Sequence 1800 BP; 747 A; 333 C; 372 G; 348 T; 0 other:

```

```

alignment_scores:
Quality: 1838.50 Length: 490
Ratio: 4.113 Gaps: 5

```


[illegible]

| | | | | |
|------|--|--------------|--------------|------|
| 943 | GGGTTATACAAAAGCGCGGTAAAG..... | 1111 | 1111 | 966 |
| 145 | lProProProAlaProleuSerGlnAlaArgInleuGlnProAlaSerArgS | 111 | ::: ::::111 | 162 |
| 967 |CCAGCTGCAAGTAAAAAT..... | CCCC | CCCC | 994 |
| 162 | erProIleleuArg.....GInGlnAAsProValAlaPheArgTyrThr | 1111::: :::: | 1111::: :::: | 176 |
| 995 | CTCCGTGTAAGAAAACAGTGAAGAAAAGCCAGTCCG.....CAG | 1111::: :::: | 1111::: :::: | 1035 |
| 177 | ValSerProThrCysSerAspThrIlyThrAlaGlySerCysSerAspSe | 1111::: :::: | 1111::: :::: | 193 |
| 1036 | GTGGAAACCCGGAAGAGACTGAGGTGAA..... | 1111::: :::: | 1111::: :::: | 1062 |
| 193 | rgIlyThrLeuSerArgleuProAlaProSerIleSerSerleuThrSera | 111 | ::: ::::111 | 210 |
| 1063 |CCTGAAGAACAAAAGATCCGAACCAAT. | 111 | ::: ::::111 | 1092 |
| 210 | snGlnSerSerleuAlaProProValSerGlyAsnProSerSerAlaPro | 111 | ::: ::::111 | 226 |
| 1093 |GGCAGTAAATCCCAAAATCCGAA | 111 | ::: ::::111 | 1116 |
| 227 | AlaProValProAlaSerIleThrAlaIleThrValSerleuValInHisSerG | 1111::: :::: | 1111::: :::: | 243 |
| 1117 | TCCTCAAGACCTCCCGCAACCCACAAACCATTA..CAAGAAAGAAAGA | 1111::: :::: | 1111::: :::: | 1163 |
| 243 | uLySerSerGlyProPheArgSerAspAsnlyuSerleuProAlaA | 1111::: :::: | 1111::: :::: | 260 |
| 1164 | TTAAATATGATTCAGACTACAGAGATACATTAACCTAAATAATGCCCGAC | 1111::: :::: | 1111::: :::: | 1213 |
| 260 | IaAlaIleuThrSerGlnHisSerIysGlyAlaSerSerIleAlaIleThr | 111 | ::: ::::111 | 276 |
| 1214 | CTGCAGCTGATTC..... | 1111::: :::: | 1111::: :::: | 1227 |
| 277 | AlaIleMetGluGlnIlyuGlyAlaProGlyThrSerProThrAsnSerI | 1111::: :::: | 1111::: :::: | 293 |
| 1227 | | 1111::: :::: | 1111::: :::: | 1227 |
| 293 | erPheValGlnInSerleuIleuHisGlyIleuIleProThrThrGlyur | 1111::: :::: | 1111::: :::: | 310 |
| 1228 |ACAAACAGTCC | 1111::: :::: | 1111::: :::: | 1237 |
| 310 | roValArgIleAsnAlaAlaArgProGlyGlyArgProGlyIlyrGlu | 1111::: :::: | 1111::: :::: | 326 |
| 1238 | CACGAAAAAGATTC.....CCAGTATATGCC | 1111::: :::: | 1111::: :::: | 1263 |
| 327 | HisSerAsnlyuSerGlyTyrIleuValSerProProGlnGlnIleArgAr | 1111::: :::: | 1111::: :::: | 343 |
| 1264 | CACAAAGATGATCCCTCAT.....CG | 1111::: :::: | 1111::: :::: | 1286 |
| 343 | rgGlyIuArgSerCysTyrArgSerIleAsnArgGlyArgInHisSerG | 1111::: :::: | 1111::: :::: | 360 |
| 1287 | ACGGGACAGGTGGATATGTTCCGATACAGATACAAAGCAC.....C | 1111::: :::: | 1111::: :::: | 1330 |
| 360 | IuArgSerGlnArgThrInSerProSerleuProAlaThrProCysPhe | 1111::: :::: | 1111::: :::: | 376 |
| 1331 | AACGTCCATCAAAATCGGAGTGTGTTAAACAGATCGAGT..... | 1111::: :::: | 1111::: :::: | 1371 |
| 377 | ValProValProProleuIlyrProProProGlnHisThrleuPr | 1111::: :::: | 1111::: :::: | 393 |
| 1372 |CTCCCTGCC | 1111::: :::: | 1111::: :::: | 1379 |
| 393 | oleuProProGlyValProProProGlnPheSerProGlnPheProSerS | 1111::: :::: | 1111::: :::: | 410 |
| 1380 | CTTGCCCATTTGACCCCTGCCCTTAACCTAACAGGCGCAATGATGCGCGAAT | 1111::: :::: | 1111::: :::: | 1428 |
| 410 | erGlnProProThrAlaGlyIlyuSerValProProPro..... | 1111::: :::: | 1111::: :::: | 422 |
| 1430 | CAGAAAGACCTGGTGGATGGAGTGGCTTAAGCCCTTAATATGCAAAATG | 1111::: :::: | 1111::: :::: | 1479 |
| 423 |GlyPheProProAlaProAlaAsnIleSerThrAlaCysPheSe | 1111::: :::: | 1111::: :::: | 437 |

| | | |
|------|---|------|
| 1480 | CAGCGGGGCCACCTCT.....ATGCATCATGTAG | 1511 |
| 437 | rProgluValProThrAlaHisSerAsnThrMetProThrThrGlnAlaP | 454 |
| | | |
| 1512 | TCACCAACATGCGACGCGGTAAACACATGCGGACGAGAGGC | 1561 |
| 454 | roLeuLeuSerAlaArgGluPheTyrArgGluGlnAsnAspLysGlyArg | 470 |
| | | |
| 1562 | CTCCCTCCAC.....TATGTCCGATACAAACCAATCCGTACAC | 1602 |
| 471 | gluSerLysPheProTyr.....SerGlySerSerTyrSerArgS | 484 |
| | | |
| 1603 | CCAATGCGTGGCGCGGTACGAGTTCGACGCGGAGGTATGATGATGATAT | 1652 |
| 484 | rSerTyrThrAspSerSerGlnGlyLeuAlaGlnHisIleHisAlaLeuT | 501 |
| | | |
| 1653 | GTCCACACCAATTTGAGTCCCAAAATTTGCTCGATATACCAAGCGGTGG | 1702 |
| 501 | hrlLeuSerProSerAlaAlaHisThrLeuAspLeuLeuHisAspHisPro | 517 |
| | | |
| 1703 | CAGCGAAGTCCGTCCGCGTCCATGTACGATCCGTGAGGCGCTTCAT | 1752 |
| 518 | HisProGluGlnGluAlaGlnAlaArgSerAlaMetIleValHisMetPr | 534 |
| | | |
| 1753 | CGCATCATGAGAGAGAGACGCGGAGAGAGCGCGTTTCGAAGTCC | 1802 |
| 534 | oAspLeuMetAspIleAlaHisAlaAlaArgSerArgSerProProTyrAla | 551 |
| | | |
| 1803 | TGAC.....CCCCACAGCTCAAGTCCCGCTAGACAA | 1837 |
| 551 | rgTyrArgArgSerArgSerArgSerProProGluPheArgGlyLysInserPro | 567 |
| | | |
| 1838 | GGCACCGCTTTAAAGCT..... | 1854 |
| 568 | ThrLysArgAsnValProArgGluGluLysGluArgGluTyrPheAsnAr | 584 |
| | | |
| 1855 |CCCATGTACGAAAGACCAACTCC..... | 1878 |
| 584 | gTyrArgGluValProProProTyrAspIleLysAlaTyrTyrGlyArgS | 601 |
| 1878 | | 1878 |
| 601 | erValAspPheArgAspProPheGluLysGluArgTyrArgGluTyrGlu | 617 |
| | | |
| 1879 |AGGATATATCTCAAGACAAAGACCGCGATCCCGGGAA | 1917 |
| 618 | ArgLysTyrArgGluTyrGlyGluLysTyrTyrLysGlyTyrAlaValAl | 634 |
| | | |
| 1918 | AGGAAGCAGAAACATAGCTACGACAGCGCATATACGC..... | 1953 |
| 634 | yAlaGlnProArgProSerAlaAsnArgGluLysPheSerProGluArgL | 651 |
| | | |
| 1954 |CACCTCGTGTAGTCGCCACCGCATATAGCTCT..... | 1989 |
| 651 | euleuProLeuAsnIleArgAsnSerProPheThrArgGlyArgArgGlu | 667 |
| | | |
| 1990 |AAGTCCCGCAGTGGCGCAATCAAAADA... | 2016 |
| 668 | AspTyrAlaAlaGlyGlnSerHisArgAsnArgAsnLeuGlyLysAsnT | 684 |
| | | |
| 2017 |TCTGACATCTCT...CGCTCGCATCTCCAAAGCC | 2048 |
| 684 | rProGluLysLeuSerThrArgAspSer...HisAsnAlaLysAspAsn | 700 |
| | | |
| 2049 | GGCGTACAAAGATGATTACAGAGACAAAGCCGTACACAAAGCCTAGTCTC | 2098 |
| 700 | roLysSerLysGluLysGluSerGluAsnValProGly..... | 712 |
| | | |
| 2099 | CCAAACAGGAGCACTTGAGCTCTCT...CCCCCGGATTCGAGCCGTTG | 2145 |
| 713 |AspLysGluLysAlaHisLysLysLysHisArgGly | 724 |
| | | |
| 2146 | CAGCTGACGATGAAGCGGTACAGAAACAGACCCGACCAAGTTCGGGA | 2195 |

| | | |
|-------|---|-------|
| 724 | sArgrArGsnGnGluGluYrGlyGluGluSerGluSerPheLeuAsnProG | 741 |
| | | |
| 2196 | AGCATTCACAAAGACCAAGGGGTATGACGAC | 2226 |
| | | |
| 741 | LuLeuLeuGluThrSerArgIysCysArgGlySerSerGlyIleAspGlu | 757 |
| | | |
| 2227 |AAGAAGAAAGGGGAAACAGGCACGAAG | 2256 |
| | | |
| 758 | ThrIysThrAspPThrLeuPheValLeuProSerArgAspAspAlaThrPr | 774 |
| | | |
| 2257 | GGCGCACGA.....AAGAGCCACAGCTCTCCACG | 2285 |
| | | |
| 774 | oValArGAspGluProMetAspAlaGluSerIleThrPheIysSerAs | 791 |
| | | |
| 2286 | CATTAGCAAGAACCGAAGCCCATATGACAGC...AATCACAGAGAGCCCGA | 2332 |
| | | |
| 791 | erAspIysAspIysArgGluYrAspIysProIysValIysSerAspIys | 807 |
| | | |
| 2333 | CTCCACCCACCAAG.....ATCACACACCGCAAAATACATCGTCGCCAG | 2376 |
| | | |
| 808 | ThrIysArgIysSerAspIysSerAlaThrAlaIysAspAsnValIle | 824 |
| | | |
| 2377 | TTCAGGCAG..... | 2385 |
| | | |
| 824 | uIysProSerIysGlyProGlnGluIysValAspIysAspArgGluYrS | 841 |
| | | |
| 2386 |CGCGAAAGTT | 2395 |
| | | |
| 841 | erProArgSerGluProProLeuIysIysAlaIysGluIuAlaThr... | 856 |
| | | |
| 2396 | CACCGAAGACG.....CCGGAAGAGACGACAGCATTTACTGACCGCG | 2439 |
| | | |
| 857 |LysIleAspSerAlaIysProSerSerSerSerGlnIysAspGl | 871 |
| | | |
| 2440 | AAGCCAGAAATTATGCGCTCCACGCCGCTCATCAACAGACAGGAATGGA | 2489 |
| | | |
| 871 | uIysValThrGlyThrProArgIysAlaHisSerIysSerAlaIysAsp | 887 |
| | | |
| 2490 | GACCAATGTGGCGAAGAGACAGACAGCCCAAGAGTCCGTGGTCAAAAGATC | 2539 |
| | | |
| 888 | ThrArgArgGluSerGlnProArgThrArgArgSerIysArgThrValPr | 904 |
| | | |
| 2540 | GCAAGAGAGAGAGAGAGACAGACAGACAGCGTGAACGCCAAG..... | 2581 |
| | | |
| 904 | oIysThrSerSerGlnIysSerGlnProValArgThrArgArgProArgS | 921 |
| | | |
| 2582 |AAACAAGAAAGACAAACG | 2600 |
| | | |
| 921 | erLeuArgIysIleAsnTyrLeuIleAlaArgGluIysAsnGluArgGlu | 937 |
| | | |
| 2601 | CGCTAGCAAG.....GAGAAAGGGATGCCAG | 2628 |
| | | |
| 938 | LysArgIysIysSerVal.....AspIysAspPheGluSerSerSe | 951 |
| | | |
| 2629 | AAGAAGAGCTCTCAGTTAATCGATGTGACTCGATATTAAACAACAGCTC | 2676 |
| | | |
| 951 | rMetIysIleSerIysValGluGlyThrGluIleValIysProSerProL | 968 |
| | | |
| 2679 | ACTAATG..... | 2685 |
| | | |
| 968 | ysArgIysMetGluIysAspValGluIysLeuGluArgThrProGluIys | 984 |
| | | |
| 2685 | | 2685 |
| | | |
| 985 | AspIysIleAlaSerSerThrThrProAlaIysIysIleIysLeuAsnAr | 1004 |
| | | |
| 2685 | | 2685 |
| | | |
| 1001 | gGluThrGlyIysIysIleGlyAsnAlaGluAsnAlaSerThrThrIysG | 1016 |
| | | |
| 2686 |AACGAGTCAAAATTATTAAG | 2700 |

```

1018 lProSerGluLysLeuGluSerThrSerSerLysIleLysGlnGluLys 1034
      ||| : : : ||| ||| : : ||| : :
2705 TATTGTCCTCCAGAGGCTCAAAAGTCCAGCATGAGATCAATGAGGCTCA 2754
      ||| : : : ||| : : ||| : :
1035 ValLysGlyLysAlaLysArgLysValAlaGlySerGlySerSerse 1051
      : : :
2755 CTT..... 2757
1051 rThrLeuValAspTyrThrSerThrSerSerThrGlySerProVala 1068
      ||||| : : :
2758 .....TCCCTTACTC 2767
1068 rGlySerGluGluLysThrAspThrLysArgThrValIleLysThrmet 1084
      : : : : : ||| : : : ||| : : :
2768 ACAACGCTACTGAAACGTTAATCCGAAAGATCCATTCCTTACTGTG 2817
      : : : : : ||| : : : ||| : : :
1085 GluGluTyrAsnAspAsnThrAlaProAlaGluAspValIleIleLeu 1101
      : : : : : ||| : : : ||| : : :
2818 GGTGCTGACGAGCATATCTTGCCCAAGAACAACTCAGCGAGGC 2867
      : : : : : ||| : : : ||| : : :
1101 tIleGlnValProGlnSerLysTrpAspLysAspAspPheGluSerGlu 1118
      : : : ||| ||| ||| ||| : : :
2868 TAATCTGTCAATCTATCCAAATGGGAAATCGACGAAATATCTTAGCTT 2917
      ||||| : : : : :
1118 lGluAspValLysThrThrGlnProIleGlnSerValGlyLysProser 1134
      : : : : : ||| : : : ||| : : :
2918 TGGAGGAT...TCCTCCAAAAAGCTGCCGGGCTCCGACGATCCGCTG 2964
      : : : : : ||| : : : ||| : : :
1135 SerIleIleLysAsnValThrThrLysProSerAlaThrAlaLysTyrTh 1151
      : : : ||| ||| : : : ||| : : :
2965 GAAATTAATCTCAGACGCTCCGCAAGGCTGAGAACGCAATATTTCGAAA 3014
      : : : : : ||| : : : ||| : : :
1151 rGluLysGluSerGluGlnProGluLysLeuGlnLysLeuProLysGlu 1168
      : : : : : ||| : : : ||| : : :
3015 GCCTATTATGCCATCAGGCTATGAGATTCAAGTTATT..... 3054
      : : : : : ||| : : : ||| : : :
1168 lAserHisGluLeuMetGlnHisGluLeuArgSerSerLysGlySerAla 1184
      : : : : : ||| : : : ||| : : :
3054 ..... 3054
1185 SerSerGluLysGlyAlaLysAspArgGlnHisSerGlySerGluLys 1201
      : : : : : ||| : : : ||| : : :
3055 ATCAATTCACAGGACAAACAGCAAGGACCGCTCCGATGTCGAGTGACAA 3104
      : : : : : ||| : : : ||| : : :
1201 sAsp.....AsnProAspLysArgLysSerGlyAlaGlnProAspL 1215
      |||| : : : : : ||| : : : ||| : : :
3105 GGAATCGCTCTCTCTCACCCAGGCGGTAAACAACG.....A 3139
      : : : : : ||| : : : ||| : : :
1215 ySGLuSerThrValAspArgLeu.....SerGluGlnGly 1226
      : : : : : ||| : : : ||| : : :
3140 GCAGGTCGGTAAAGATAGGCTGGGCACCAAGTTTCCAAATGATAGAAC 3189
      : : : : : ||| : : : ||| : : :
1227 HisPheLysThrLeuSerGlnSerSerLysGluThrArgThrSerGluLys 1243
      : : : : : ||| : : : ||| : : :
3190 CGTTCGCGAGACAAGTCGAAGGCGAGCGCGCGCAAGAGCTCCGA 3239
      : : : : : ||| : : : ||| : : :
1243 sHisGluSerValArgGlySerSerLysAspPheThrProGlyAlaArg 1260
      : : : : : ||| : : : ||| : : :
3240 CGACGATGGCAACCGCGAGCTCGATGTCATGGCAGCGGAGAGAGG 3289
      : : : : : ||| : : : ||| : : :
1260 sPlysLysValAspTyrAspSerArgAspTyrSerSer...SerLysArg 1275
      ||| : : : ||| : : : ||| : : :
3290 ACAACAGA.....TCCCGGACACAGCGCGCGCTTACAGAGAG 3327
      : : : : : ||| : : : ||| : : :
1276 ArgAspGluArgGlyLysLeuAlaArgArgLysAspSerProArgLys 1292
      : : : : : ||| : : : ||| : : :
3328 AGCAGAGAGAGCT.....TCGTACAAGCGCAAGCTCCGCG..... 3360
      : : : : : ||| : : : ||| : : :
1292 yLysGluSerLeuSerGlyLysLysSerLysLeuArgGluGluAlaArgAspL 1309
      : : : : : ||| : : : ||| : : :
3361 .....GAGGACGACAAAGCTGAGGCGCGCAAGAC.... 3387
      : : : : : ||| : : : ||| : : :
1309 euProLysLysGlyAlaGluSerLys.....LysSerAsnSer 1321

```

```

3388 .....AAGAACAGTCCGAATCCAAACGACGAAAGCATGATCAAAACAT 3432
      ||| : : : ||| ||| : : ||| : :
1322 SerProProArgAspLysLysProHisAspHisLysAlaProTyrGluTh 1338
      ||| : : : ||| : : ||| : :
3433 AGCAGACACTCGGATCGAGGCGGCGCAAAACACCAAGTCCAGCGAG 3482
      : : : : : ||| : : : ||| : : :
1338 rLys.....ArgProCysG 1343
      : : : : : ||| : : : ||| : : :
3483 CCGAGTGGTCTCTCTGTAAACAGCGTGGTCTCTCCCAAACTCTGTC 3532
      : : : : : ||| : : : ||| : : :
1343 lGluThrLysPro.....ValAspLysAsnSerGlyLysGlu 1355
      : : : ||| ||| : : : ||| : : :
3533 GTCCACACAAACCGCTCCGCAAGTTGCTGACACCGAGTTGCTGACACAG 3582
      : : : : : ||| : : : ||| : : :
1356 Arg.....GluLysHisAlaAlaGluAlaArg 1364
      : : : : : ||| : : : ||| : : :
3583 TTAGTTGTAAATATGATTAACAGATACAGAAAGAGGCGGCTCGCA 3632
      : : : : : ||| : : : ||| : : :
1364 gAsnGlyLysGlu 1368
      ||||| |||
3633 CAACGGCATGAG 3645
      : : : : : ||| : : : ||| : : :
seq_name: /SID51/gcdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL05842
seq_documentation_block:
ID ABL05842 standard; cDNA: 7385 BP.
XX
AC ABL05842;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 12008.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PI 11-JUL-2000; 2000US-0614150.
XX
PA (PERE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WP1: 2001-656860/75.
XX
P-PSDB; ABB61739.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 12008; 21pp + Sequence Listing; English.
XX
PS
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB5737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WFO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX

```

SO Sequence 7385 BP; 2213 A; 1708 C; 1645 G; 1819 T; 0 other;

alignment_scores: length: 1513
 quality: 530.50 caps: 67
 ratio: 0.792
 percent similarity: 44.283 percent identity: 20.952

alignment_block:
 us-09-811-045a-1 x ABL05842 ..

Align seg 1/1 to: ABL05842 from: 1 to: 7385

```

35 LysGluLysProProPheLeuProGluGlu.....Pr 45
   |||||
2594 CCGGAGAACGATGCGCTGAGAACGATTCAGCCGATTTGTGCTGCC 2643
   |||||
45 OsSerSerSerGluGluAspAspProIleProAlaGluLeuLeuCysL 62
   |||||
2644 TGCTGTACAAACCAAGAG.....ATACCGGAGGATCTGATATGCC 2684
   |||||
62 euIleCysIlyAspIleMetThrAspAlaValAlleProCysCysGly 78
   |||||
2685 GCATATCCGAGATATATTCGTGATGCTGATGATACCTGCTGGCGA 2734
   |||||
79 AsnSerSerCysAspGlyCysIleArgThrThreLeuLeuSerAsp 95
   |||||
2735 AGTTCCTTTGTGACGACTGTGCGAACCTCCTATTGAGATCAGAGA 2784
   |||||
95 sHisThrCysProThrCysHisGlnAsnAspValSerProAspAla 112
   |||||
2785 TAGTGAAGTCCCGCATTCGCAAGAGAAAGACTGTCCTGCTGCTCG 2834
   |||||
112 leaIleAsnLysPheLeuArgGlnAlaValAsnAsnPhelysAsn 128
   |||||
2835 TACCTATTCGGTCTTTCAGGAATTCGCTGAACGCTTAAAAATGAG 2884
   |||||
129 GlyTyrThrLysArgLeuArgGlyGlnLeuPro..... 139
   |||||
2885 GGGTATTAACAAGACGCGGCTAAG.....CCAGGTAAAGTTCAAT 2928
   |||||
140 .....ProPheLeuPheLeuValProp 147
   |||||
2929 CGTTTCATGCGCTTAATGCAAAATGATGAAATTTCTCTTTTAC 2977
   |||||
147 ropProArgProLeuSerGlnArgAsnLeuGlnProArgSerArg 163
   |||||
2978 .....GTAAAAATGAGAAAAACCTCCT 3001
   |||||
164 IleLeuArg.....GlnGlnAspProValAlaPheArgTyrThr 178
   |||||
3002 GTTCAAAAAGAAAGTGAAGAAAAACCCAGTCCG.....GAGGTG 3042
   |||||
178 rProThrCysSerAspThrLysThrAlaGlySerCysSerAspSer 195
   |||||
3043 ACCCGAAGAGCTGAGTGAAA..... 3064
   |||||
195 hrLeuSerArgLeuProAlaProSerIleSerSerLeuThrSerAsn 211
   |||||
3065 .....CTTCAAAAAGCAAAAAGAAATCCGAAACCAAT..... 3094
   |||||
212 SerSerLeuAlaProProValSerGlyAsnProSerSerAlaPro 228
   |||||
3095 .....GGCAGTAAATCCGCCAAAATCGGAATCTCC 3123
   |||||
228 oValProAspIleThrAlaThrValSerIleSerValHisSerGly 245
   |||||
3124 AGACCTCTCCCGCAACACAGAACCATCA...CAGAGAGAGAAAGT 3170
   |||||
245 eAspGlyProPheArgSerSerAspAsnLysLeu.....ProAla 259
   |||||
3171 ATGATTCAGACTACGAGATTAACATTACCAATAAAATGCCCGAGC 3220

```

```

260 AlaAlaLeuThrSerGluHisSerLysGlyAlaSerSer..... 272
   |||||
3221 GCTGATTCACAAAGTCCGCAAGAAAGTGAGACATACATATTTCAC 3270
   |||||
273 .....IleAlaIleThrAlaLeuMetGluGluLysGlyValP 285
   |||||
3271 TCTGCACATAGATACATTAATTCCTTAATGTTTCAGAGATCC.... 3316
   |||||
285 roGlyThrSerProThrAsnSerIlePheValGlyGlnSerLeuHis 301
   |||||
3316 ..... 3316
   |||||
302 GlnGlnLeuIleProThrThrGlyProValArgIleAsnAlaAla 318
   |||||
3316 ..... 3316
   |||||
318 oGlyGlyGlyArgProGlyTyrProLysHisSerAsnLysLeuGly 335
   |||||
3317 .....CCGAGTATTCACACAGAAAGTGAATCCTTCAT.... 3349
   |||||
335 alSerProProGlnGlnIleArgArgGlyGluArgSerCysTyrArg 351
   |||||
3350 .....CGAAGGACAGGTCGATTAATGTTTCC 3376
   |||||
352 IleAsnArgGlyArgHisHisSerGluArgSerGlnArg...Thr 367
   |||||
3377 .....GATCAGATCACAAACAGCACCAAGTCATCAATAATC 3411
   |||||
367 rProSerLeuProAlaThrProCysPheValProValProProPro 384
   |||||
3412 GGAGTCTGTAAACAGAGATCGAGTCTCTGCTGCTGCTGCTGCTG 3461
   |||||
384 eu..... 384
   |||||
3462 TGCCTACCTCCAGGCGCATATGATGCCGATCAAGAAAGTCCGCA 3511
   |||||
385 .....TyrProProProProHisThrLeuProLeuProProGly 358
   |||||
3512 TCGAGTCCCTAATAAGCCCT.....TATATGCAATGACAGCGG 3555
   |||||
398 lProPro.....ProGlnPheSerProGlnP 407
   |||||
3556 ACCTCCTATGCACATGATGATGATGATGATGATGATGATGATG 3605
   |||||
407 heProSer.....SerGlnProProThrAlaGlyTyrSerValPro 421
   |||||
3606 TTAACAACATGAGGACAGAGGCTCCCTCAGATGAGTTCTTTCC 3655
   |||||
422 ProGlyPheProProAlaProAlaAsnIleSerThrAlaCysPhe 438
   |||||
3656 GTCCGAGATCCG..... 3667
   |||||
438 oGlyValProThrAlaHisSerAsnThrMetProThrThrGlnAla 455
   |||||
3667 ..... 3667
   |||||
455 euLeuSerArgGluGluPheTyrArgGluGlnAsnAspLysGlyArg 471
   |||||
3667 ..... 3667
   |||||
472 SerLysPheProTyrSerGlySerSerTyrSerArgSerSerTyr 488
   |||||
3668 .....CTACCATGTGATGATGATGATGATGATGATGATGATG 3703
   |||||
488 pSerSerGlnGlyLeuAlaGlnHisIleHisAlaLeuThrLeuSer 505
   |||||
3704 .GAACCTCGAGGTGTGACGATCCATACGACGCGGCTGCTGATCT 3752
   |||||
505 eAlaAlaHisThrLeuAspLeuLeuHisAspHisProHisProPro 521
   |||||
3753 CAATA.....TTTAAAGTTCCAAAGAGACTATACACACACATG 3796
   |||||
522 GluAlaGluAlaArgSerAlaMet.....IleVal 531

```

```

3797 AAGAGTAAATTAAGCTGCGCTGCGCCCTTATGCTTTGTTGAATAT 3846
      : : : : : ||| : : : : :
531 LHisMetProaspLeuMetAspIleAlaHisIaIarSer..... 544
      : : : : : ||| : : : : :
3847 GGTATTCCTAATTAATAACAGTATGCTTACATACGTTAAGCTTTT 3896
      : : : : : ||| : : : : :
545 .....ArgSerProProTyrArgArgTyrArgSerArg 555
      : : : : : ||| : : : : :
3897 GAATGTATTAATTAATCTTAATAATTCAGCTATGTCGCCGATCAAAACCA 3946
      : : : : : ||| : : : : :
556 SerArgSerProProGluPheArgGlyGlnSerProThrLysArgAsnVa 572
      : : : : : ||| : : : : :
3947 TCCGTACACCA..... 3958
      : : : : : ||| : : : : :
572 LProArgGluGluLysGluArgGluTyrPheAsnArgTyrArgGluValP 589
      : : : : : ||| : : : : :
3959 .....ATGC 3962
      : : : : : ||| : : : : :
589 roProProTyrAspIleLysAlaTyrTyrGlyArgSerValAspPheArg 605
      : : : : : ||| : : : : :
3963 GTGCCCGCTACGGA...TCTGCAGGCGGAGGTATGATATGATATGTC 4009
      : : : : : ||| : : : : :
606 AspProPheGluLysGluArg.....Tyr.ArgGluTyrGluVal 618
      : : : : : ||| : : : : :
4010 CAACCATTTCACTCCCAAAATTTAGCTCGATATACCAAGGCGTGACAGC 4059
      : : : : : ||| : : : : :
618 rGlys.....TyrArgGluTyrPyr 624
      : : : : : ||| : : : : :
4060 GAAGGTGCGGTTCCGGGTGAGTATCTTTAGGCGGTTGTAAGACCGGCGAT 4109
      : : : : : ||| : : : : :
625 GluLysTyrTyrLysGlyTyrAlaValGlyAlaGlnProArgProSerAl 641
      : : : : : ||| : : : : :
4110 GACAAAGTATCCCAATGTTTGCAGTC..... 4136
      : : : : : ||| : : : : :
641 aAsnArgGluAspPheSerProGluArgLeuLeuProLeuAsnIleArg 658
      : : : : : ||| : : : : :
4137 .....CATTGACGATCCGTT 4152
      : : : : : ||| : : : : :
658 snSerProPhe.....ThrArgGlyArgArg...GluVal 668
      : : : : : ||| : : : : :
4153 GAGGCGCTTCAATCCCATCATGACAGAGAGAGCGGAAGAGGTGGAC 4202
      : : : : : ||| : : : : :
668 sPtyrAlaAlaGlyLysSerHisArgAsnArgAsnLeuGlyLysAsnTyr 684
      : : : : : ||| : : : : :
4203 GCTTCGAAGCTCGACCGCCACAGGTCAAGTCC..... 4237
      : : : : : ||| : : : : :
685 ProGluLysLeuSerThrArg.....AspSerHisAs 695
      : : : : : ||| : : : : :
4238 CCGGATAGCAAAAGCACCCTTTAAGTCTCCCATGTACGAAAGACAA 4287
      : : : : : ||| : : : : :
695 nAlaLysAspAsnProLysSerLysGluLysGluSerGluAsnValProG 712
      : : : : : ||| : : : : :
4288 CTCACGGGATATCTCAAGAGACAAAAGACCGGATCCCGGAAAGAGAAC 4337
      : : : : : ||| : : : : :
712 LysAspGlyLysGlyLysAsnLysHisLysLysHis...ArgLysArgArgAsn 727
      : : : : : ||| : : : : :
4338 GAGAACATAGCTACGACGCGCATATACGCCACCTCTCTAGTGGCAG 4387
      : : : : : ||| : : : : :
728 GluGluLysGlyGluGluSerGluSer..... 736
      : : : : : ||| : : : : :
4388 CCGAATGATGCTCTAAGTCCCGAGGTGACAGAAATCAAAAGGTAGTTT 4437
      : : : : : ||| : : : : :
737 .....PheL 738
      : : : : : ||| : : : : :
4438 CCACCAAGCCCTTTCCATGCTGTACATTCATGCTACTCTCTCTTTT 4487
      : : : : : ||| : : : : :
738 euAsnProGluLeuLeuGluThrSerArgLysCysArgLysSerSerGly 754
      : : : : : ||| : : : : :
4488 TAGATCTGGACATCGTCGCTCGATCTCCAAAGCGGCTACAGAGAG 4537
      : : : : : ||| : : : : :
755 IleAspGluThrLysThrAspThrLeuPheValLeuProSerArgAspAs 771
      : : : : : ||| : : : : :
4538 ATTACAGAGACAAAGCCGTACAAAGACCTAGTCTCCCAAAAGGAG... 4584
      : : : : : ||| : : : : :
771 palAthrProValArgAspGluProMetAspAlaGluSerIleThrPheL 788
      : : : : : ||| : : : : :
4585 .....GCAGTTGAGCCTCTCTCCCGGATTCGAGCGCTTG..... 4620
      : : : : : ||| : : : : :
788 ySerValSerAspLysAspLysArgGluLysAspLysProLysValLys 804
      : : : : : ||| : : : : :
4621 ..CAGCTACGAGTGAAGACGGCTACAGSAAACACACCGCACAGTTG 4668
      : : : : : ||| : : : : :
805 SerAspLysThrLysArgLysSerAspGlySerAlaThrAlaLysLysAs 821
      : : : : : ||| : : : : :
4669 GAAGCATCAAAAGCAGCAGCGGTGATGACAGC.....AAGAAGAG 4709
      : : : : : ||| : : : : :
821 pasnValLeuLysProSerLysGlyProGluGluLysValAspGlyAsp 838
      : : : : : ||| : : : : :
4710 AGGGAAAAACAGCACAGAGAGCGCCACGAAAGAG.....CACA 4750
      : : : : : ||| : : : : :
838 rGluLysSerProArgSerGluProProLeuLysLysAlaLysGluGlu 854
      : : : : : ||| : : : : :
4751 GGTCTCGAGCATTTAGCAAGAACCG..... 4776
      : : : : : ||| : : : : :
855 AlaThrLysIleAspSerValLysProSerSerSerSerGluLysAsp.. 870
      : : : : : ||| : : : : :
4777 .....AAGCGAATGACACCACTACAGAGAGCT 4805
      : : : : : ||| : : : : :
871 .....GluLysValThrGlyThrProArgLysAlaHisSerLys 884
      : : : : : ||| : : : : :
4806 GACTCCACCACGCAAAAGATCAC...ACACGAAATGACT.....G 4843
      : : : : : ||| : : : : :
884 eAlaLysAspThrArgArgLysSerGlnProArgThrArgArgSerLys 900
      : : : : : ||| : : : : :
4844 CTGCCAGATTGAGCGCAACGCAAGTTTCACGACAGCGCGGAAAAAGAT 4893
      : : : : : ||| : : : : :
901 Arg.....ThrValProLysThrSerSerGluLysSerGluProVal 914
      : : : : : ||| : : : : :
4894 CACGACGATTATCTGACCGCAAGGCCAGAAATATGCTCCACCCCT 4943
      : : : : : ||| : : : : :
914 L.....ArgThrArgArgProArgSerLeuArgLysIleAsnTyrLeuI 929
      : : : : : ||| : : : : :
4944 CATCAACGACACGGAATGAGAACCAATGTGGCAAGGAAACAAAG... 4989
      : : : : : ||| : : : : :
929 leAlaArgGluLysAsnGluArgGluLysArgLysLysSerValAspLys 945
      : : : : : ||| : : : : :
4990 ..GCCAAGAGTCCGTTGTCAAAGATCCGCAAGAGAAAGAGACAG 5037
      : : : : : ||| : : : : :
946 AspPheGluSerSerSerMetLysIleSerLysValGluGlyThrGluI 962
      : : : : : ||| : : : : :
5038 GAC.....AAGCGTACGCGCAAGAAAAA 5060
      : : : : : ||| : : : : :
962 eValLysProSerProLysArgLysMetGluGlyAspValGluLysLeu 979
      : : : : : ||| : : : : :
5061 CAAGAAAGACAAAGCCGCTAAGAGAGAAAGGGATCCCAAG... 5106
      : : : : : ||| : : : : :
979 luArgThrProGluLysAspLysIleAlaSerSerThrThrProAlaLys 995
      : : : : : ||| : : : : :
5107 .....AAG 5109
      : : : : : ||| : : : : :
996 LysIleLysLeuAsnArgGluThrGlyLysLysIleGlyAsnAlaGlu.. 1011
      : : : : : ||| : : : : :
5110 AGCTCTCTCAGTTAATCGA...TCTGACTCGGATTAATTAACAACGCTCACT 5156
      : : : : : ||| : : : : :
1012 ...AsnAlaSerThrThrLysGluProSerGluLysLeuGluSerThs 1027
      : : : : : ||| : : : : :
5157 AATGAACGAGTCAAAATTAAGATATGCTCTCCAGGCGTCAAGTCCA 5206
      : : : : : ||| : : : : :
5207 GCATTGAGATCAATGCGGCTCAACTT..... 5232
      : : : : : ||| : : : : :
1044 AlaGlySerGluGlySerSerSerThrLeuValAspTyrThrSerThse 1060
      : : : : : ||| : : : : :
5232 ..... 5232

```

```

1060 rSerThrGlySerProValArgLysSerGluGluLysThrAspThrL 1077
      |||||..... |||||.....
5233 .....TCCCTACTCAACAGCGTACGAAAAACGTATATCCGA 5269
1077 ysaGythrValIleLysThrMetGluGluTyrAsnAspAspThrAla 1093
      |||||..... |||||.....
5270 AGAGTCAATCCATCTTACGTGGTGGCTGCTACGACGATATCTTTCG 5319
1094 ProLagLysAspValIleIleMetIleGlnValProGlnSerLysTrpAs 1110
      |||||..... |||||.....
5320 CCAAGAACCAATCAGCAGGAGGCTAATCTGTCAATCTATCTCAATGGGA 5369
1110 pLysAspAspPheGluSerGluGluGluAspValLysThrThrGlnProI 1127
      |||||..... |||||.....
5370 AATCGACGAGATATCTTAGTGTGGAGAT...TCTCCAAAAAAGCTG 5416
1127 IeGlnSerValGlyProSerSerIleIleLysAsnValThrThrLys 1143
      |||||..... |||||.....
5417 CCGGGGCGCTCCGACGATCCGTCGGAATATCTCAGACGCTCGCAAG 5466
1144 ProSerAlaThrAlaLysTyrThrGluLysGluSerGluGlnProGluL 1160
      |||||..... |||||.....
5467 GCTGAGAACGCAATATTTCGAAGGCTATTATGCATCAGGCGCTATGCA 5516
1160 sLeuGlnLysLeuProLysGluAlaSerHisGluLeuMetGlnHisGluL 1177
      |||||..... |||||.....
5517 GTTCAAGTATAT..... 5529
1177 euArgSerSerLysGlySerAlaSerSerGluLysGlyArgAlaLysAsp 1193
      |||||..... |||||.....
5530 .....ATCAATTCACAGACACACAGCAGGAC 5556
1194 ArgGlnHisSerGlySerGluLysAsp.....AsnProAspLysAr 1207
      |||||..... |||||.....
5557 CGCTCCGAGTTCGAAGTACAGAGATCGCTCCTCCTCACCAGCGATA 5606
1207 gLysSerGlyValGlnProAspLysGluSerThrValAspArgLeu... 1222
      |||||..... |||||.....
5607 CAACAGC.....ACAGGTCGGTAAAGATAGCGTGGGCA 5641
1223 .....SerGluGlnGlyHisPheLysThrLeuSerGlnSerSer 1235
      |||||..... |||||.....
5642 CCAAGATTTCCAATGATAGAAGCCGTCGCGACCAAGTCGAAGGCGCG 5691
1236 LysGluThrArgThrSerGluLysHisGlnSerValArgGlySerSerAs 1252
      |||||..... |||||.....
5692 CGCCGGGCGCGAAGAGCTCCGACGACGATCGAACCGCGCAGGTGCGA 5741
1252 nLysAspPheThrProGlyArgAspLysLysValAspTyrAspSerArg 1269
      |||||..... |||||.....
5742 TCGTCATGGCGCGCGAAGAGGAGCAACAGA.....TCCGCGC 5779
1269 sPTyrSerSer...SerLysArgArgAspGluArgGlyGluLeuAlaArg 1284
      |||||..... |||||.....
5780 ACAGGGGCGCGCTTCAGAGAGAGGAGGAGCGT.....TCCGAC 5820
1285 ArgLysAspSerProProArgGlyLysGlnSerLeuSerGlyGlnLysSe 1301
      |||||..... |||||.....
5821 AAGGAAAGCTGCGCG.....GAGGACGA 5843
1301 tLysLeuArgGluGluArgAspLeuProLysLysGlyAlaGlnSerLys. 1317
      |||||..... |||||.....
5844 CAACGTGAGGGCGCGAAG.....AAGGACAGCGCAATCCAAAGC 5884
1318 .....LysSerAsnSerSerProProArgAspLysLysProHis 1330
      |||||..... |||||.....
5885 ACGGAAAGCATGATCAAAACATAGCGACGATCGGAGGCGCGCC 5934
1331 AspHisLysAlaProTyrGluThrLys..... 1339
      |||||..... |||||.....
5935 AAAAACACCAAGTCCAGCAGCGCGAGTGTCTCTCTGTGAACGCCGT 5984

```

```

1340 .....ArgProCysGluGluThrLysPro.....V 1348
      |||||..... |||||.....
5985 GGTGTCCTCCCAACCGCTGCTCCAGACAAACCGCTCCGCAAGTTCG 6034
1348 alaAspLysAsnSerGlyLysGluArg..... 1356
      |||||..... |||||.....
6035 TCGACACCGAGTTCGTGAGCAGCTTACTGTAATATGATACAGATA 6084
1357 GluLysHisAlaAlaGluAlaArgAsnGlyLysGlu 1368
      |||||..... |||||.....
6085 CAGAAAGAGGCGCGCTCCTCGACACAGCGCATGGAG 6120
seq_name: /st01/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.AA182535
seq_documentation_block:
ID   AA182535 standard; cDNA; 439 BP.
XX
AC   AA182535;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polynucleotide seq ID NO 2595.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PE   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
XX
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Dmanac RT;
XX
DR   WPI; 2001-514838/56.
XX
DR   P-PSDB; AAO02604.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 2595; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AA179941-AA193841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_ptc_sequences.
XX
SQ   Sequence 439 BP; 154 A; 88 C; 88 G; 109 T; 0 other;

```

```

alignment_scores:
Quality: 528.00      Length: 144
Ratio: 4.157        Gaps: 0

```


Percent Similarity: 88.194 Percent Identity: 71.528

alignment_block:
US-09-811-045a-1 x AA182535 ..

Align seg 1/1 to: AA182535 from: 1 to: 439

```

1041 ARGVSLVALAGLSERGLUGLSERSESRTHLEUVALASPTYRTH 1057
      |||||||.....:|||||.....:|||||.....:|||||
7  CAAAGAGCTGAGCTGAGAGATGCTACAGTGTCTGATTAC 56
1057 rserthrserserthrglyserprovalarglysergluluglyst 1074
      |||||||.....:|||||.....:|||||.....:|||||
57  CATTCCGACTCACTGAGAGCATGCTGTGCGGAATCTGAAGACCA 106
1074 hAspThrlysarThrVal11elysThmetgluglulTrysAsnAsp 1090
      |||||||.....:|||||.....:|||||.....:|||||
107  CAGATACTAAGCGAACTGTGATTAAACATGAGACATATATATGAC 156
1091 AsnThrAlaProAlaGluAspVal11lelleMetlleGlnValProGlnse 1107
      |||||||.....:|||||.....:|||||.....:|||||
157  AATACCGCTCCAGCTGAAAGATGTTATCATTTAGATCAGGTTCTCAATC 206
1107 rlystTPAspLysAspAspPhegluSerglugluluglulAspVallystThrT 1124
      |||||||.....:|||||.....:|||||.....:|||||
207  CATATGGAATCAAGATGACTTGAATCTGAAGACCAAGATGATAATCCA 256
1124 hGlnProIleGlnSerVal1GlyLysProSerSerlelleLysAsnVal 1140
      |||||||.....:|||||.....:|||||.....:|||||
257  CACAGCCTATATCAAGTGTAGGAAACCTGCTACTTATATAAAATGTT 306
1141 ThrThrLysProSerAlaThrAlaLysTYrThrGlnLysGlnSerGlnG 1157
      |||||||.....:|||||.....:|||||.....:|||||
307  CGTACAAAGCCATCTCTCTATGTCAGATATCCTGAGAAAGAAAGTGAAGCC 356
1157 nProGluLysLeuGlnLysLeuProLysGluAlaSerHISGluLeuMetG 1174
      |||||||.....:|||||.....:|||||.....:|||||
357  ATTCCGCTAAATTCGAAATTCACCATGAGCGTCAATGAAATGATCATAAC 406
1174 lnhISGluLeuArgSerSerLysGlySerAla 1184
      |||||||.....:|||||.....:|||||.....:|||||
407  CTCATGAGGTTAAAGTTCATATACTCTGCA 438
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:AAV21511
seq_documentation_block:
ID  AAV21511 standard; DNA; 6755 BP.
AC  AAV21511;
XX
DT  17-AUG-1998 (first entry)
XX
DE  Staphylococcal bacteriocin BacRI operon.
XX
KW  BacRI; bacteriocin; antimicrobial; antibacterial; antibiotic;
KW  Moraxella bovis; infectious bovine keratoconjunctivitis; cancer;
KW  therapy; ds.
XX
OS  Staphylococcus aureus strain UT0007 (ATCC 55800).
XX
PN  W09812319-A1.
XX
PD  26-MAR-1998.
XX
PE  18-SEP-1997; 97WO-US16758.
XX
PR  17-SEP-1997; 97US-0931999.
PR  19-SEP-1996; 96US-0710561.
XX
PA  (UNIV ) UNIV KANSAS STATE RES FOUND.
XX
PI  Crupper SS, Iandolo JF,
XX

```

DR WPI: 1998-230316/20.

XX Therapeutic proteinaceous substances from *Staphylococcus aureus* -
PT useful to inhibit growth of wide range of prokaryotic or eukaryotic
PT cells, e.g. *Moraxella bovis* causing infectious bovine
PT keratoconjunctivitis
XX
PS Claim 2; Page 19-23; 38pp; English.

XX This polynucleotide comprises the bacteriocin BacRI operon of
CC *Staphylococcus aureus* UT0007. The sequence of the BacRI operon
CC was determined by N-terminal sequencing of purified BacRI peptide
CC (see AA54171), with back-translation and plasmid analysis. The
CC BacRI operon includes the BacRI gene (see AAV21510), a homologue of
CC the *cylM* gene of the cytolysin operon of *Enterococcus faecalis*
CC whose function is involved in the maturation of pre-cytolysin,
CC an ATP-transporter gene, *bio1* and *bio2* genes related to
CC lactococcal biosynthesis and modification, and a gene involved in
CC immunity function. BacRI peptides can be produced by construction
CC of an expression vector containing an oligonucleotide or operon
CC coding for BacRI, and use of the vector to transform host cells for
CC BacRI expression. The entire BacRI operon has been cloned into
CC plasmid pUB110, and *Bacillus subtilis* transformants secreted the
CC recombinant BacRI peptide into the medium. Bacteriocin BacRI is
CC active against many Gram-positive and Gram-negative organisms such
CC as *Bordetella bronchiseptica*, *Pasteurella multocida* and
CC *Staphylococcus aureus*; *Moraxella bovis*, causing infectious bovine
CC keratoconjunctivitis, is especially sensitive. BacRI can also be
CC used as an anti-cancer agent.
XX

Sequence 6755 BP; 3903 A; 1252 C; 1600 G; 0 U; 0 other;

alignment_scores:
Quality: 477.00 Length: 995
Ratio: 0.958 Gaps: 40
Percent Similarity: 50.050 Percent Identity: 21.508

alignment_block:
US-09-811-045a-1 x AAV21511 ..

Align seg 1/1 to: AAV21511 from: 1 to: 6755

```

550 ARGArgTYrArgSerArgSerArgSerProProGluPheArgGlyGlnse 566
      ||||| |||:|||||:|||||
2890 CGAGCGCAACGCAAAAGAAACAGA..... 2913
566 rProThrLysArgAsnValProArgGlnGluLysGluArgGluTYrPhea 583
      :|||:||||| |||||: |||:|||||:
2914 .....AGAAAGAAC...CCGAGGAGAACGAAAGACGACAAAGCAGCA 2953
583 snArgTYrArgGluValProProTYrAspIleLysAlaTYrTYrGly 599
      :|||
2954 GGAGA..... 2958
600 ArgSerValAspPheArgAspPheGlnLysGluArgTYrArgGluTr 616
      |||||:|||||:|||||:|||||:
2959 .....GAGAAAAAGACAGACAGACAC 2981
616 pGluArgLysTYrArgGluTrTYrGluLysTYrTYrLysGlyTYrAlav 633
      :|||:||||| |||||: |||||:
2982 CAGAAAGCCGACGCAAAAGAAAGAAAAA.....GAGGGACACACAG 3025
633 aLgLYaLagInProArgProSerAlaAsnArgGlnAspPheSerPro... 648
      :|||:||||| :|||: |||||: |||||:
3026 CAGGGCGACCGGACAGAAACACAGAAACAGAAACCGGACCGAGG 3075
649 .....GluArgLeuLeuProLeuAs 655
3076 CAGCAAAAGGGGAGAGAAAAACAAAAACGAAAGACACAAAGACAC 3125
655 nIleArgAsnSerProPheThrArgGlyArgArg..... 666
      :|||:|||||: |||||:

```

```

3126 CACAGAAACCAACCCAAAGACAGAGAAAGAGGGGCAACACAGAAA 3175
667 .....GluAspTyr 669
3176 CGAAAAAACAAACGGCCACAGCCAAACAAAGCCGCAAGGAAAGACAGA 3225
670 AlaAlaGlyInSerHisArgAsnArgAsnLeuGly ..... 681
3226 GCGAAAGGGCAGACAGAGGACAAACGACAAAGAAACAAAGCAAAAGCCGC 3275
682 .....GlyAsnTyrPro GlyLysLeuSerThrArgAspSer 693
3276 AACCCCAAAAGAGAGAGACCCAGAAAGGGCGACACACAGAGAGAGAA 3325
694 HisAsnAlaLys...AspAsnProLysSerLysGlyLysGlySerGluAs 709
3326 CACGAGAAAGAAAGCCAAACGAAAGCCACGACGAAAGCAGAGAAAAAG 3375
709 nValProGlyAspGlyLysGlyLysHisLysLysHisArgLysArg 726
3376 GAGGACAGAGAAAGAAAGGCGCCACAAAGCAAGAAAAAGGAAAGAAA 3425
726 rgsAsnGluGluLys.....GlyGluGluSerGluSerPheLeuAsn 739
3426 ACAAGAGCAAAAGAGAGGAGAGAGAAAGAAAGAAAGACACAAACAA 3475
740 ProGluLeuLeuGluThrSerArgLysCysArgLysSerGlyLys 756
3476 ACCGAGCAGAAAGAAAGAAAGGAAAGAAAGAAAGAAAGAAAGAAACA 3525
756 pGluThrLysThrAspThrLeuPheValLeuProSerThrArgAspAla 773
3526 GAAACACAAAGACAGAGA..... 3541
773 hrProValArgAspGluProMetAspAlaGluSerIleThrPheLysSer 789
3542 ..GAAAAAAAGAAAGAAACACAGACAAACAGAAACAAAGCGAAACAC 3589
790 ValSerAspLysAspLysArgGluLysAspLysPro..... 801
3590 GAAAAAACGAGACAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3639
802 .....LysValLysSerAspLysThrLysArgLys 812
3640 CGACCAAAAGAAAGAAAGCCGAAACCAAGAAAGAAAGAAAGAGAGGA 3689
812 eraSpGlySerAlaThrAlaLysLysAspAsnValLeuLysProSerLys 828
3690 AAGAAAAAGAAAGAAAGCGAGCAAAA.....GCCGACGAAAGAAC 3727
829 GlyProGluGluLysValAspGlyAspArgGluLysSerProArgSerG 845
3728 GGAGGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCGCGGA 3777
845 uProProLeuLysLysAlaLysGluGluAlaThrLysIleAspSerVal 862
3778 CAGACCAACAAAGAAAGGAGCAGAGAGAAAGCAAAAGAAAGAAAGAAC 3827
862 ysProSerSerSer.....SerGlnLysAspGluLysValThrGly 875
3828 AAAAGAGAGAGCAGAAAGAAAGAAAGAGAGAGCAAAAGAAAGCCGCG 3877
876 ThrProArg.....LysAlaHisSe 882
3878 ACAAGAGAAAGAGAGAGAGAGCAAGCAAGAAAGCAAAAGCAAAAGAG 3927
882 rLysSerAlaLysAspThrArgArgLys..... 891
3928 CAAGACCAAAAGGAGCGCAAGCGCAGGAGCAGACAGAGAGAGAGAAAC 3977
892 .....SerGlnProArgThrArgArgSerLys 900
3978 AAAACGAAACCAAAACGAGAGAAACAGCCAGCCAGAAAGAAAGAAAG 4027

```

```

901 ArgThrValProLysThrSerSerGlnLysSerGluProValArgThrArg 917
4028 GAACACCGGAAAAAACCAAGCAGACACAAACCAAAAGAAAGAAAGACAG 4077
917 gArg.....ProArgSerLeuArgLysIleAsnTyrLeu 929
4078 AAAGAGGCGCAAAACCCACCCCAAAAGAAAGAAAG..... 4114
929 leaAlaArgGluLysAsnGluArgGluLysArgLysLysSerValAspLys 945
4115 ..GCCGCAAAAGAGCAGCAGAGCAAGAAAGCCCAAGAAAA..... 4150
946 AspPheGluSerSerMetLysIleSerLysValGluGluThrGlu 962
4151 .....GACCGGAACCAAAAGCAAAAGCAAGACAGAGAGAGCAAGC 4191
962 eValLysProSerProLysArgLysMetGluGlyAspValGluLysLeu 979
4192 AACCCACCCACAA..CGACAGAAAGCAAAAGGA.....AAAGGGCGG 4232
979 LuArgThrProGluLysAspLysIleAlaSerSerThrThrProAlaLys 995
4233 AAAACGACCAAAAGAAAGAAAGCAAA.....AACACCAAGAGAGAAA 4273
996 LysIleLysLeuAsnArgGluThrGlyLysLysIleGlyAsnAlaGlu 1012
4274 AAG.....GCCACCGACAAACACCGCCAGCAAAAGAAC 4305
1012 nAlaSerThrThrLysGluProSerGluLysLeuGluSerThrSerSer 1029
4306 CCAGAAACAAAGAAAGAAAGAAAGCAAGAAAG.....AAGCACAAAGCC 4349
1029 ystiLysGluGluLysValLysGlyLysAlaLysArgLysValAlaGly 1045
4350 AACAAAGAAAGAAAGAAAGCCGAAAGAAAGAAAGAAAGAAAGCAAAA 4399
1046 SerGluLysSerSerSerThrLeuValAspTyrThrSerThrSerThr 1062
4400 CAAAGAGAAACCCCAACAAAC.....AGCAAAAGAAAGCAAGAGAGAC 4443
1062 rGlyLysSerProValArgLysSerGluGluLysThrAspThrLysArg 1079
4444 AGGAAAGCGCAGAACCCAGAGAAAGAAAGGAAAGAAAGCAAAAGAAAG 4493
1079 hrValIleLysThrMetGluGlu..TyrAsnAsnAspAsnThrAlaPro 1095
4494 GAAAGCAAAAGCAAAAGCAGAGAGCAACACAAAGAAAGCGCGAGAC 4543
1095 aGluAspValIleIleMetIleGlnValProGlnSerLysTrpAspLys 1112
4544 GGAAGAA.....AAACACCCCGAAAGAAAGAAAGCAAAAGAAAG 4578
1112 spAspPheGlu..SerGluGluGluAspValLysThrThrGlnProIleG 1128
4579 ACAACACAAAGCAAAAGCCCAAGAGAGAAAGAAAGCAACCAAGCA 4620
1128 nSerValGlyLysProSerSerIleIleLysAsnValThrLysPro 1145
4621 .....AAAAAGAAAGCGGGAAGAGAG 4642
1145 eraAlaThrAlaLysTyrThrGlu..... 1152
4643 AGGAAGACCGCCAGAAAGCAAGCAACAGAAAGCGAGGGAGCCAGAAC 4692
1153 LysGluSerGluGluProGluLysLeuGluLysLeuProLysGluAlaSe 1169
4693 AAAAGAGAGCAAAAGCAGAGAGAGCAAAAGCAAAAGAGAAAGAAAGA 4742
1169 rHisGluLeuMetGlnHis..... 1175
4743 ACACAAAGAAAGAAAGAAAGCAAGAAAGCAAAAGCAAAAGCAAGACAGA 4792

```



```

1 MetMetGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsnTh 17
  |||
312 ATGATGGAAGTGAAGATCTTAATATGAAGGTGCAATGCTTACCAACAC 361
  |||
17 rGlyLysrYrAlaIleProThrIleAspAlaGluAlaIleGlyL 34
  |||
362 TGGAAAATGTCATACCAATACCAATATGATGACAGACATATGCAATGGGA 411
  |||
34 ySLysGluLysProPheLeuProGluGluProSerSerSerSerGlu 50
  |||
412 AGAAAAGAGAAACCTCCCTCTTACAGAGAGACCATCTTCTTCACAGA 461
  |||
51 GluAspAspProIleProAlaGluLeuLeuCysLeuIleCysLysAsp 67
  |||
463 GAGATGATGATCCATCCAGATGATGATGATGATGATGATGATGAT 511
  |||
67 eMethrAspAlaValIleProCysCysGlyAsnSerSerCysAspG 84
  |||
512 TATGACTGATGCTGTGATGATGATGATGATGATGATGATGATGATG 561
  |||
84 LuCysIleArgThrThrLeuLeuGlu 92
  |||
562 AATGTAAGAAAGTGCATCTTGGAA 587
  |||

```

seq_name: /SIS1/9cgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF28296

seq_documentation_block:

ID AAF28296 standard; RNA; 2565 BP.

AC AAF28296;

DT 05-APR-2001 (first entry)

XX LPL mRNA.

KW Lipoprotein lipase; LPL; S447X; cardiovascular; diabetes; obesity;

KW deficiency; ss.

OS Homo sapiens.

PN W0200100220-A2.

PD 04-JAN-2001.

PE 23-JUN-2000; 2000WO-CA00762.

PR 24-JUN-1999; 99EP-0202048.

XX (UYBR-) UNIV BRITISH COLUMBIA.

PA (AMST-) AMSTERDAM MOLECULAR THERAPEUTICS BV.

PI (UYAM-) UNIV AMSTERDAM ACADEMIC HOSPITAL.

DR WPI; 2001-112388/12.

PT Use of lipoprotein lipase S447X therapeutic for the preparation of a
 PT pharmaceutical composition for the treatment of coronary heart disease,
 PT atherosclerosis, angina pectoris, hypertension, diabetes, cachexia and
 PT obesity

PS Claim 5; Fig 4; 51pp; English.

XX The present invention relates to use of a lipoprotein lipase (LPL)
 CC S447X therapeutic for the preparation of a pharmaceutical
 CC for the treatment of an LPL-responsive condition. Diseases that may be
 CC treated are complete LPL deficiency, chylomicronemia, hyperlipidemia,
 CC partial LPL deficiency, pancreatitis, hypertriglyceridemia,
 CC hypobetalipoproteinemia (low high density lipoprotein-cholesterol),
 CC cardiovascular disease, coronary heart disease, coronary artery
 CC disease, atherosclerosis, angina pectoris, hypertension,
 CC cerebrovascular disease, coronary restenosis, peripheral vascular
 CC disease, diabetes, cachexia or obesity.

```

xx
SQ Sequence 2565 BP; 1020 A; 739 C; 806 G; 0 U; 0 other;

alignment_scores:
  Quality: 413.50      Length: 937
  Ratio: 0.884        Gaps: 50
  Percent Similarity: 49.947      Percent Identity: 25.080

alignment_block:
US-09-811-045A-1 x AAF28296 ..

Align seg 1/1 to: AAF28296 from: 1 to: 2565

517 ProHisProGluGluAlaGluAlaArgSerAlaMetIleValHisMe 533
  2 CCCCCCCCCCAAGGAAAGC.....
533 rProAspLeuMetAspIleAlaHisAlaArgSerArgSerProProTyra 550
  23 .....GCCACACGCGCGCCGACCCCAAA 47
550 rGArgTyArgSerArgSerArgSerProProGluPheArgGluGlu.. 565
  48 GGGCGACGCCAGCGCCAAACCGCGCGCCAGCCCGCCAGCGCGCGCGC 97
566 .....SerProThrLysArgAsnValProArgGluG 576
  98 GGGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGAG 147
576 LuLysGluArgGluTyrrPheAsnArgTyrrArgGluValProProProTy 592
  148 AGCAAAAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG... 194
593 AspIleLysAlaIleTyrrGlyArgSerValAspPheArgAspProPheG 609
  195 .....CGAGCGCGCGCGCGCGCGCGCGCGCAAAAGA..... 221
609 uLysGluArgTyrrArgGluTyrrPheLysArgLysArgLysArgLysArg 626
  222 .AGAGACGACAC.....GAAAGAAAGCCCAAGAGACCCGAAAGACA 261
626 ySTyrrTyrrLysGlyTyrrAlaValIleGlnProArgProSerLysAsn 642
  262 CAGCGAGACAGCGCACCCAGCGAGAGAGAGAGAGAGAGAGAGAGAG 311
643 .....ArgGluAspPheSerProGluArgLeu 652
312 ACAGCAGCAAAACCAAGGACCGAGCGAGAGAGAGAGAGAGAGAGAGG 361
652 uProLeuAsnIleArgAsnSerProPheThrArgGlyArgArgGluAsp 669
  362 GCCAAACGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 411
669 yTAla.....AlaGlyGlnSerHisArgAsnArgAsnLeu 680
  412 GGGCAGCGCGCGAGCAACCGCGCGCGCGCGCGCGCGCGCGAGCA 461
681 Gly.....GlyAsnTyrrProGluLysLeuSerThrArgAspSerHisAs 695
  462 GGAAGGCGCGCGAGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAG 511
695 nAlaLysAspAsnProLysSerLysGluLysGluSerGluAsnValPro 711
  512 GGAACAGCGCGAGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAG 549
712 GlyAspGlyLysGlyAsnLysHisLysLysHisArgLysArgArgAsnG 728
  550 .....AAAGAAAGCAACAGAAAGCGCGCGA 575
728 uGluLysGlyGluGluSerGluSerPheLeuAsnProGluLeuLeuGluT 745
  576 CCAGCGAGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 625

```

| | | |
|--------|---|------|
| 1027 | rsertLysIleLysGlnGlnIulysValLysGlyLysAlaLysArgLysValA | 1044 |
| 1028 | | 1045 |
| 1029 | | 1046 |
| 1030 | | 1047 |
| 1031 | | 1048 |
| 1032 | | 1049 |
| 1033 | | 1050 |
| 1034 | | 1051 |
| 1035 | | 1052 |
| 1036 | | 1053 |
| 1037 | | 1054 |
| 1038 | | 1055 |
| 1039 | | 1056 |
| 1040 | | 1057 |
| 1041 | | 1058 |
| 1042 | | 1059 |
| 1043 | | 1060 |
| 1044 | | 1061 |
| 1045 | | 1062 |
| 1046 | | 1063 |
| 1047 | | 1064 |
| 1048 | | 1065 |
| 1049 | | 1066 |
| 1050 | | 1067 |
| 1051 | | 1068 |
| 1052 | | 1069 |
| 1053 | | 1070 |
| 1054 | | 1071 |
| 1055 | | 1072 |
| 1056 | | 1073 |
| 1057 | | 1074 |
| 1058 | | 1075 |
| 1059 | | 1076 |
| 1060 | | 1077 |
| 1061 | | 1078 |
| 1062 | | 1079 |
| 1063 | | 1080 |
| 1064 | | 1081 |
| 1065 | | 1082 |
| 1066 | | 1083 |
| 1067 | | 1084 |
| 1068 | | 1085 |
| 1069 | | 1086 |
| 1070 | | 1087 |
| 1071 | | 1088 |
| 1072 | | 1089 |
| 1073 | | 1090 |
| 1074 | | 1091 |
| 1075 | | 1092 |
| 1076 | | 1093 |
| 1077 | | 1094 |
| 1078 | | 1095 |
| 1079 | | 1096 |
| 1080 | | 1097 |
| 1081 | | 1098 |
| 1082 | | 1099 |
| 1083 | | 1100 |
| 1084 | | 1101 |
| 1085 | | 1102 |
| 1086 | | 1103 |
| 1087 | | 1104 |
| 1088 | | 1105 |
| 1089 | | 1106 |
| 1090 | | 1107 |
| 1091 | | 1108 |
| 1092 | | 1109 |
| 1093 | | 1110 |
| 1094 | | 1111 |
| 1095 | | 1112 |
| 1096 | | 1113 |
| 1097 | | 1114 |
| 1098 | | 1115 |
| 1099 | | 1116 |
| 1100 | | 1117 |
| 1101 | | 1118 |
| 1102 | | 1119 |
| 1103 | | 1120 |
| 1104 | | 1121 |
| 1105 | | 1122 |
| 1106 | | 1123 |
| 1107 | | 1124 |
| 1108 | | 1125 |
| 1109 | | 1126 |
| 1110 | | 1127 |
| 1111 | | 1128 |
| 1112 | | 1129 |
| 1113 | | 1130 |
| 1114 | | 1131 |
| 1115 | | 1132 |
| 1116 | | 1133 |
| 1117 | | 1134 |
| 1118 | | 1135 |
| 1119 | | 1136 |
| 1120 | | 1137 |
| 1121 | | 1138 |
| 1122 | | 1139 |
| 1123 | | 1140 |
| 1124 | | 1141 |
| 1125 | | 1142 |
| 1126 | | 1143 |
| 1127 | | 1144 |
| 1128 | | 1145 |
| 1129 | | 1146 |
| 1130 | | 1147 |
| 1131 | | 1148 |
| 1132 | | 1149 |
| 1133 | | 1150 |
| 1134 | | 1151 |
| 1135 | | 1152 |
| 1136 | | 1153 |
| 1137 | | 1154 |
| 1138 | | 1155 |
| 1139 | | 1156 |
| 1140 | | 1157 |
| 1141 | | 1158 |
| 1142 | | 1159 |
| 1143 | | 1160 |
| 1144 | | 1161 |
| 1145 | | 1162 |
| 1146 | | 1163 |
| 1147 | | 1164 |
| 1148 | | 1165 |
| 1149 | | 1166 |
| 1150 | | 1167 |
| 1151 | | 1168 |
| 1152 | | 1169 |
| 1153 | | 1170 |
| 1154 | | 1171 |
| 1155 | | 1172 |
| 1156 | | 1173 |
| 1157 | | 1174 |
| 1158 | | 1175 |
| 1159 | | 1176 |
| 1160 | | 1177 |
| 1161 | | 1178 |
| 1162</ | | |

```

      ...      |||      :|||      |||      |||
2240 CGGCCAAAAAGAGCAAGCAAGCAACACAGG.....CGAAGAACCC 2283
1310 OLYSLYSGLYAlaGluSer...LYSLYSerAsnSerSerProPArgA 1326
      |||||      |||:      |||:      |||:      |||:
2284 AAAAAAAGAGAGAGGAGGACCCCAACAAACGAGGGGAGCCAGG 2333
1326 SPlySLYSPROHISAsPHISLYSAlaProTYrGLuThrLYSArgProCys 1342
      |||      |||      |||      |||
2334 GCAAA.....CAAAGACCAAAAGCAGC 2356
1343 GLuGLuThrLYSPROVALAsPlyAsnSerGLYSLYGLuArgLYSHI 1359
      :|||      |||||      :|||      |||||
2357 ACAAGCAGGGAACCGGCCGAAAAACGAAACAGAGCGG.....CA 2400
1359 SAAlaGLAlaArgAsnGLYSLYGLuSerGLYAlaAsnCYSHIS 1375
      |||      |||      |||      |||
2401 CAAAGCAAAACAGGAGGAGGAAAGAGCAGCAAGAAAGCAGGAGC 2449
seq_name: /SIDSL/gc9data/geneseq/geneseqn-emb1/NA2001B.DAT.ABL05592

```

seq_documentation_block:

```

ID ABL05592 standard; cDNA; 18506 BP.
XX
AC ABL05592;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11258.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
   pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN MO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001, 2001MO-US09231.
XX
PR 23-MAR-2000, 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
P-PSDB; ABB61489.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
   genes from Drosophila and for elucidating cell signalling and cell-cell
   interactions -
XX
PS Claim 1; SEQ ID NO 11258; 21bp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
   capable of detecting 1000 or more genes from Drosophila. The invention is
   useful in developmental biology and in elucidating cell signalling and
   cell-cell interactions in higher eukaryotes for the development of
   insecticides, therapeutics and pharmaceutical drugs. The invention
   discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
   sequences (AB101840-AB16175) and the encoded proteins
   (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
   specification, but was obtained in electronic format directly from WIPO
   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 18506 BP; 5394 A; 4676 C; 4890 G; 3546 T; 0 other;

```

alignment_scores:

```

Quality: 401.00      Length: 1772
Ratio: 0.505      Gaps: 82
Percent Similarity: 44.808      Percent Identity: 20.203

```

alignment_block:

US-09-811-045A-1 x ABL05592 ..

Align seq 1/1 to: ABL05592 from: 1 to: 18506

```

37 LysProProPhe.....LeuProGLuGLuProSerSer.....Se 48
      :|||      |||||      :|||      |||||
3619 CGGCCCGCAGGAGCACCCTTGCCCAAAATGCCAAAGCTTACGAGC 3668
48 rSerGLuGLuAspAspProLeuProAlaGLuLeuLeuCYSLIECYSL 65
      :|||      |||||      :|||      |||||
3669 CGGCCCAACCAAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 3718
65 ySAspILeMetThrAspAlaValAlaValleProCYSLYSLYSLY 81
      :|||      |||||      :|||      |||||
3719 GAGGCCCAACACCGAAGGTGCTCGAGTCCAAAGCAGCAGCGCTCG 3768
82 CysAspGLuCYSLIEArgThrThrLeuLeuGLuSerAspLYSHISThrCY 98
      |||      |||      |||      |||
3769 TGC.....AGGCCACAGC 3782
98 sProThrCysHISGLInAsnAspValSerProAspAlaLeuIleAlaSNL 115
      :|||      |||      |||      |||
3783 ACCGTCCTAGCGGAGTACAGCAGTCCCA..... 3813
115 ySPHeuArgGLInAlaValaSNAsnPhelysAsnGLuThrGLYThr 131
      |||      |||||      :|||      |||||
3814 .....GTGAAAGCGAGTCCAAACAGCAGG 3837
132 LysArgLeuArgLYSLYGLuLeuProProPheLeuPheLeuValProPro 148
      :|||      |||||      :|||      |||||
3838 CGGAAGCAAGACAGCGGCTA.....CTGGGCGCACCCAA 3872
148 oArgProLeuSerGLInArg...AsnLeuGLInProArgSerArgSerPro 164
      |||||      :|||      |||||      |||||
3873 GCGACCCAGAGAAACCAATCTCAAGAGAGACTGCTGCGCTTACACCTC 3922
164 eLeuArgGLInAspProValAlaPheArgTYrThrValSerProThrc 181
      |||      |||      |||      |||
3923 GAAGCGCGCACAGACCCGGAAGTCGTAAGCAGCAGCAAGCGAAG 3972
181 ySserAspThrLYSHISArgLYSLYSerSerAspSerGLYThrLeuSer 197
      :|||      |||      |||      |||
3973 CGCGAGCTTAATAAAAGACA...CGTTTGACAAGGAGGAGCAC.... 4013
198 ArgLeuProAlaProSerILeSerSerLeuThrSerAsnGLInSerSer 214
      :|||      |||||      :|||      |||||
4014 .....ACGACCTAGTCT 4027
214 uAlaProProValSerGLYAsnProSerSerAlaProAlaProAlaPro 231
      :|||      |||||      :|||      |||||
4028 GGTGCCACCCCTCTGCAGATGAGCAGACGCTCCAAAGAACTGCAAG 4077
231 sPILeThrAlaThrValSerILeSerValHISerGLuLYSAspLY 247
      :|||      |||||      :|||      |||||
4078 ACTTAACCGCTTCTCGAGAACTCGACGAGAGAACACCGTGAATGAGC 4127
248 ProPheArgAspSerAspAsnLYSLYLeuProAlaAlaLeuThrSe 264
      :|||      |||||      :|||      |||||
4128 GACCTTAAGGAGGACAGAGGTGTACGGGAATGAAAGCGGTGTTAG 4177
264 r...GLuHISerLYSGLYAlaSerSerLeuAlaIleThrAlaLeuMet 280
      :|||      |||      |||      |||
4178 TCGTATGAATGAAGAACGACAGCAGCACCAACAGATCAAGCGCGCTTC 4227
280 LuGLuLYSGLYAlaProGLYThrSer..... 288
      :|||      |||||      :|||      |||||
4228 GAGAAATGCTGCGAGAGAACCGGCGATGTGAAATGAGCTGATGAA 4277

```

```
289 .....ProTpaSerIlePheV 295
4278 GAGGAGAGTACTGATCATCGAAAAGAGAGATGGAGCATATACCGA 4327
295 aIeGIInSerIeuleuHISgIyInIeulPro..... 306
4328 AGATTCGATCGTTGACAGAGAGACGATGACCAAGAAAGAGAGATCC 4377
307 ...ThThGIyProValArgIleAsnAlaAlaArgProGIyGIyAlar 332
4378 AGAAGACACGAGCGTACTCGACAGATCGAAGAAAGCAAGAAAGAGAG 4427
322 gProGIyTrpGIuHISerAsnIyLeuGIyTrIeulValSerProProg 339
4428 GCCGAG..... 4434
339 IngIInIeArgArgGIyIuArgSerCysTrArg.....Ser 351
4435 .....AGGAGATCGAGCGCGCATTCGCAAGGTAGAAAGCCGTCGAGC 4476
352 IleAsnArgGIyArgHISerISerGIuArgSerGIuArgThGIInSerP 368
4477 GTAAACGGCGACTAGAGGGTGCCTCTCGAGCGAGATGAGTCGAACTG 4526
368 oSerIeulProAlaThr.....ProCysPheValProV 379
4527 GACGTGACCGCGAGACGACAAAGATMAAGCTGAGTGCAGATATCAT 4576
379 aIProProProIeulTyTrProProProHISerIeulProIeulPro 395
4577 TGCCTACTGCCAAGGATATTG.....CCAAGTCACGCA...CCGAGAGC 4617
396 ProGIyValProProGIuHISerISerProGIuHISerProSerGIInP 412
4618 AGTTGGCGAAGCCCGCGAGAGAGAGCTCTCAT.....CGCCAGCC 4658
412 oProThrIaGly.....Tys 418
4659 CCCGAGAGAGAGCTAGCAAAAAGACTTCGATAGACAGAGCATCAAT 4708
418 eValProProProGIyPheProProAlaProAlaAsnIleSerThrAla 434
4709 CGAGAGCCCGCTCGACGTTCTACCG.....TCA 4737
435 CysPheSerProGIyValProThrIaHISerAsnThrMetProThrTh 451
4738 ATCTTCAAGAAAGCTAACCG...AAGAGAGTTCTCGGCTACCATTTAG 4784
451 rGIuAlaProIeulSerArgGIuInIuPheTy..... 462
4785 TCGGGGGCCACCACTGCTCGGAC.....TACCCGAGATGAGCGCAT 4828
463 .....ArgGIuGIaAsn.AspIyGIyArgIuSerIy 473
4829 TCGGTGATTCAGATTAAGAGAGACCTCGTATCGAGAGAAATATGTGA 4878
473 sPheProTyTrSerGIySerSerTyTrSerIuArgSerTyTr...ThAsp 489
4879 AAGAGAGACCAAGAGAGCGGAGCGCATGTATGGCCACCGTACGACT 4928
489 eSerGIIn.....GIyIeulAGIInIS.IleHISAlaIeulTh 501
4929 CTTCGCCGAGAGAGCTCATTTGCAATCGATACCATCTTCCCTCAGCTAC 4978
501 rIeulSerProSerAlaAlaHISerIeulAspIeulHISAspIleProH 518
4979 CAAGGATGCCCCCAAGAGCGCAACGCTTAGCGGAGACTCCCG 5028
518 IsPheProGIuGIuAla.....GIuAlaArgSer 527
5029 ACTCAGAGAGAGGGTGCCTGCCATGAGATGACATTTAGGCGCCAGAG 5078
528 AlaMet.....IleValHISMetProAspIeulMetAspIleAlaH 541
5079 AACCTTCCCGCATGTCAATCAAAAGCCGTGAC.....GAATGTGACGA 5122
541 sAla.ArgSerArgSerProProTyTrArgTrArgTrArgSerArgSer... 556
5123 TTTAACCGCTTCATAGAGAAAGCCGATCTTGCGCTATACGAAAGACAGC 5172
557 .....ArgSerProProGIuP 562
5173 AGAGCGCCGCGCCAAAAGTATCTCCATTAAGAGAGAGTCGCCAAAAGAA 5222
562 eArgGIyInSerProThrIySArgAsnValProArgGIuGIuIySGIuA 579
5223 GAAAAGAGAGCGACGACGAAAGAGAGATA...AGGTAGAGAGATCGCA 5269
579 rGIuTyTrPheAsnArgTyTrArgGIuValProProProTyTrAspIleIy 595
5270 ATGGGGGATGAACCGAATTAAG.....TCGACATCTCTC 5304
596 AlaTyTrTyTrGIyArgSerValAspPheArgAspProPheGIuIySGIuAr 612
5305 ACSTACTACTCAAGAGT.....CG 5324
612 gTyTrArgGIuTyTrGIuArgIySArgTrArgIuTrP.....T 624
5325 GTACAGAGAGTGGCCGAGAAAG.....TCGTGTTATTCGAACTACAGT 5368
624 yGIuIySArgTyTrIySArgIyTyTrAlaValaIeGIuInProArgProSer 640
5369 GCGAAGAAACAGACAGAGAGATGTGGAGGCCACTACAGTCATTAACCAAG 5418
641 AlaAsnArgIu..... 644
5419 AAAAAAGAGAGATCTAATGAGACAGGTTAAGATTAAGAGAGCATGA 5468
645 .....AspPheSerProGIuArgIeul..... 652
5469 CAAAAATAGAGTCTGGAATCATTTACGAAAGAGCAAGCTAATAAGTCGGC 5518
652 .....P 653
5519 TAGCACTCCCGAGAAAGAGACAGACAGACATTACATCGATGAGAGC 5568
652 ..... 652
5569 TTCCTGCCCACTGGCGAGCCGACCACTGTGCCACCAATACGCCAAA 5618
653 .....P 653
5619 GACCGGAGATACCGGCTCCATTGAAGTCCCCCTACCATTTGAGGAGGC 5668
653 rIeulAsnIleArgAsnSerProPheThrArgGIyArgArgGIuAspTyTr 669
5669 CATTTAGGTGAGGTGACAGCGCAAGCAAGAACCCCAAAAACGAGTCCAG 5718
670 AlaAlaGIyInSerHISArg.....AsnArgAsnIeulGIyIyAsnTy 684
5719 CTCGGAGAGAGGCTATCAAGACGAAAGTGCCTTGCCTCAAGAGAG 5768
684 rProGIuIyIeulSerThrArgAspSerHISAsnAlaIySAsp..... 698
5769 ACCTCCGAGACCTGAATCTGGACAGAGAGCGTCAAGGAAGACACTGACGA 5818
699 .....AsnPro.LysSer.....L 703
5819 GACGAATATCCAAAGAAATCCCACTACCTCCGACCTGAATCTGAGGACA 5868
703 ySGIuIySGIuSer.....GIuAsn 709
5869 AAGATTAAGAGTCTCCATTCGCTTCAGAGAGAGCTTCTGACCGGAAATC 5918
710 ValProGIyAspGIyIySGIuAsnIySISIySISIySISArgIySArgTr 726
```



```

1184 ..... 1184
7730 AAAATCTCTTGGCCCTCCAGAAAGCTTCCAGACCAGCTTCGCTTGACG 7779
1185 ..... SerSerGluLysGlyArgAlaLysAspArgLHis 1196
7780 AAAGTATTAAAGGACGACAGCTGAGAAAGTCTAAGGAAGAACTGAGCGACAG 7829
1197 SerGlySerGlu..... 1200
7830 TCGGTGGCCGCAAAATCTCTTGGCCCTCCAGAAAGCTTCCAGACCAG 7879
1201 ..... LysAspAsnProAspLys..... 1206
7880 ATCCGCTGCAGAAAGTATTAAAGCGAAAGCTGAGAGCTTAAGGAAGAA 7929
1207 ..... ArgLysSerGlyAlaGln..... ProAspLysGluSer 1217
7930 CTAGCGGAGAGTGGCGCGCGCAAAATCTCTTGGCCCTCCAGAAAGCT 7979
1218 Thr..... ValAspArgLe 1222
7980 TCCAGACCAGCTTCGCTTGCCAGAAAGTGTAAAGATGAAGCTGATTAAGTC 8029
1222 userGluGlnGlyHisPheLysThrLeuSerGlnSerLysGlu.... 1237
8030 CAAGAAGAGATCTAGCGGGAAATCGATGGCCGAAGCGGTAAGCCGACGA 8079
1238 ..... ThrArgThrSerGlu 1242
8080 GTATCAAGGATCATCAATCTCTCTGAAGGAAGTTTACAGACCAAGATCT 8129
1243 LysHisGluSerValArgGlySerSerAsnLysAspPheThrProGlyArg 1259
8130 GTGGCGGAGAGGTGTAAACACATCTCTGCAATCCAGAGAACCATCCCG 8179
1259 GasPylsLysValAspTyrAspSerLysAspTyrLysSerSerLysArg 1276
8180 TCGAATAATCTGTA..... GCAGGAAGTGTCAAGCGGAAATGTGCTA 8220
1276 rGAspGluArgGlyGluLeuAlaArgArgLysAspSerProProArgGly 1292
8221 GAGAGACCAATCTCTTGGAAATCCAGAGGACGCTCAAGGCCA..... 8264
1293 LysGluSerLysSerGlyGlnLysSerLysLeuArgGluGluArgAspLe 1309
8265 ...GAATCAGTG..... GTTGACAGCGTTAAGACGAACTGTGAATA 8302
1309 uProLysLysGlyAlaGluSerLysLysSerAsnSerSerProProArg 1326
8303 GCMAAAGATCGTCGCAATTAACAAAGAAAGTGTATACCTCCAAAG 8352
1326 sPlysLysProHisAspHisLysAlaProTyrGluThrLysAspProCys 1342
8353 CTAAG..... GATGATAAGTCTCCAAAGAAAGTTTATCAACACGTG 8393
1343 GluGluThrLysProValAspLysAsnSerGlyLysGluArgGluLysHis 1359
8394 TGTATGACGAGACATATTAGGAAGACGCGACCAACCCATGAAGCATC 8443
1359 sAlaAlaGluAlaArgAsnGlyLysGluSerSerGlyAlaAsnCysHis 1375
8444 ACAACCGCAAGAGTCGAAGGAATCGATTGTGTAACCATTAAGAGCTTCA 8493
1376 ValTyrLeuThrArgGlnThrLeuProTyrArgSerTyrPheLeuGlu 1392
8494 GTCTTAAGGACGAGA...AGTCTCAATTGGCTTGAAGAAAGACGCTTAGG 8540
1392 YArg 1393
8541 CCAG 8544
seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAH817955

```

```

seq_documentation_block:
ID AAH81795 standard; DNA: 9027 BP.
XX
XX AAH81795:
XX
XX 21-SEP-2001 (first entry)
XX
DE Human differential transcription-associated cDNA SEQ ID 304.
XX
XX Differential transcription; human; rat; tumour cell; cytostatic;
XX Ras modulator; Class II tumour suppressor gene; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX W0200157058-A2.
XX
XX 09-AUG-2001.
XX
XX 31-JAN-2001; 2001WO-EP01003.
XX
XX 31-JAN-2000; 2000DE-1004102.
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;
XX Grips M, Hellriegel M, Schmitt A, Sers C;
XX WPI: 2001-483415/52.
XX
XX Nucleic acids differentially expressed between tumor and normal cells,
XX useful for diagnosis or therapy of tumors and for screening active
XX agents -
XX
XX Disclosure: Page 457-459; 579pp; German.
XX
XX This invention describes a nucleic acid (I) with differential expression
XX between tumour and normal cells and which has cyrostatic activity. (I)
XX work as modulators of Ras actively by inducing expression of tumour
XX suppressor genes. (I), and polypeptides encoded by them, are useful as
XX targets for diagnosis or therapy and in screening to determine the
XX effects of an active compound (potential pharmaceutical) on a cell line,
XX particularly for diagnosis and treatment of tumors, especially by
XX modulating expression of (I) (by gene therapy, antisense RNA or ribozyme
XX methods) or by modulating the amount and/or location of (I)-encoded
XX polypeptides (by administration of the polypeptide or its activator,
XX antibody (optionally as a conjugate) or inhibitor). The method allows
XX identification of many Class II tumour suppressor genes (i.e. genes that
XX are not primary targets for tumour-initiating mutations).
XX AAH81492-AAH82376 represent the human and rat derived nucleic acid
XX fragments described in the method of the invention.
XX
XX Sequence 9027 BP; 2206 A; 2831 C; 2223 G; 1767 T; 0 other;
XX
XX
XX alignment_scores:
XX      Quality: 400.50      Length: 1784
XX      Ratio: 0.515      Gaps: 77
XX Percent Similarity: 43.610      Percent Identity: 20.684
XX
XX alignment_block:
XX US-09-811-045A-1 x AAH81795 ..
XX
XX Align seg 1/1 to: AAH81795 from: 1 to: 9027
XX
XX 31 AAtailegLyLSuLSuLySProPhoPheLeuProGluGluProSerSe 47
XX |||::: |||::: |||::: |||::: |||::: |||::: |||:::
XX 367 GCCTGTGTAAGCGGGCTATCTCTGACATCTCGAGCACGACGACCAAGC 416
XX 47 rtsrsergluLgLuAsPAsPProileProAlaIleuLeuLeuScyLeuIleC 64
XX
XX 417 GCGC.....GTGCGAGCTGCGAGC 442

```

[illegible]

1114 TCCCTTCCTCCAGG.....CGA 1131
1115
317 ArgProGlyGlyGlyAlaArgProGlyTyrProAlaHisSerAsnLysLeuGlyTyr 333
1116
1132 CGCGGGAGGAGATGCGCCCTTTCAGTAGAACCCAGGTACTACACGACACACA 1181
333 PLeuValSerProProGlnGlnIleArgArgGlyLysLysSerCysTyrA 350
1133
1182 ACGGCGTAGTAGCCCGAGACTGCTACAGAAACAGGCTAGCAGCCCTTAG 1231
350 rGSerLLeaSnArgLysArgHisHisSerGlnArgSerGlnArgHisGln 366
1232 AAGCAAAAGATAAAGACAAAGAGAGAAATCGCAACTGCACCTAGCCCC 1281
367 SerProSerLeuProAlaThrProCysPheValProValProProPro 383
1282 TCTCCGAAAGAGACGACGACAGCGCCACAGACCACTGCTGCCACTCGGT 1331
383 OLeu.....TyrProProProHisThrLeuProL 394
1332 CTTTCGTAGGAGACATGGCGGCTCCACACACCCCTGCAGCACCCCTT 1381
394 eu.....ProProGlyValProProProGlnPheSer..... 404
1382 TAAGCCAGAGGACGACTGAACCCCATCTGAGAGCCCTCCAACTCGGAC 1431
405ProGlnPheProSerSerGlnProProThrAlaGlyTyrSe 418
1432 CGTTACCAACCTTAAGTCTCCGAGAAACTCCCACTCTTCTTCGTCAG 1481
418 rValProProProGlyPheProProAlaProAlaAsnLysSerThr...A 434
1482 GAGCAGCCACCATCC.....CCTCAACCTACCAAAAGTTTCTCGGCATG 1525
434 lAcysPheSerProGlyValPro..... 441
1526 CCAGCTTTCCTCCAGAAAGTCTTAACCTGCTCCAGCTCCAGGCTCCAC 1575
441 441
1576 CGAGAGATTCTCTCTCCACATCTAAGAAATCGCTACATGCGCGAGC 1625
442ThrAlaHisSerAsnThr..... 447
1626 AAACGGGATAAATACATTCATTCATACCCCTCCCGTAGATGGGAGGT 1675
448MetPro 449
1676 CCGGTAGCCCTGCGACCGGTAAAGAGAGGCGGATCTCGGTCCGAACCCCT 1725
450 ThrThrGlnAlaProLeuLeuSerArgGlnGlnPheThrArgGlnGlnAs 466
1726 ACCAAGAGAGGTCTATTCTGCATCCGATCCCACTGCCAGTGGGTAGGTCA 1775
466 nAspLys.....GlyArgLysSerLysPheProTyrSerLysers 480
1776 GCTGTGCACAGAGTGGGAGAACTCTGAACCCCAAGCAGAGTGGCGGT 1825
480 er.....TyrSerArgSerSerTyrThrAsp...Se 489
1826 CTAGGTCTCTCAGCACCGCTGGTGTGAGGACAAATAATACCCAGAGA 1875
489 rSerGlnGlyLeuAlaGlnHisLysHis...AlaLeuThrLeuSerPro. 504
1876 AAGAGCAGGTCTAGGTACGACAAAGCGAGGAGGTCCCACTAGATATCCC 1925
505SerAlaAla 507
1926 AGCCACTAGGCGGTAGATCTCGTTCTAGAACACAGGCCCGGGGAGGT 1975
508 HisThrLeuAspLeuLeuHisAspHisProHisProGlnGlnAlaGln 524
.....


```

986  silealaserthrthrproalalylysilelysleuasn..... 1000
    |||:|||||: ||| :||| |||:|||||:
3751  ....GCTACGTGCATCACCTCCTCTAGACAGAAAGCAAAATTTAGTCCCTTTC 3796
1001  .....Arg 1001
3797  CAGTACAGAGATAGCCCTGAGTCTTCACTGTATTCAAGACACACTTGA 3846
1002  gluthrlylslysileglyasnalaaglnasnaIasethrthrlysgl 1018
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
3847  ACCCGCCCAAGAAAGAGAGTGTCTGTGTCATCTCCAGAAACAAAGA 3896
1018  uProserglulysleugluserthrSerSerlysllelysglnlulysv 1035
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
3897  GCAAAATATGTCATTTGCTTACGTCACGCAAGATGAAGATTAAATGAGG 3946
1035  aLlysGlyLysAlaLysArgLysValAlaGlySerGlnLysSerSer 1051
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
3947  TGGTAGAGAGATCTGAGAACCCGACGCCAAATCTGTCATTTGCT 3996
1052  ThrleuValAspTyrThrSerThrSerThrThrGlySerProvalAr 1068
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
3997  TCAGAACTTAAGAAATGTCACCAAGTACTTGAATCATCTCCT.... 4041
1068  glySerSerGlnLulysThrAspThrLysArgThrValIleLysThrMetg 1085
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4042  .....GAAGTAGAAGAAAGCCCTGCTGTCTTGCACCTTTC 4078
1085  lugalutyrAsnAsnAspAspThrAlaProAlaGlnAspValIleIleMet 1101
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4079  ATCAGAGCCAGTCAACAGGCTTCTTTGGAGCA..... 4110
1102  lleGlnValProGln.....SerLysTrpAspLysAspAspPheGln 1116
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4111  GTAGAGATCCCTCAATGCTCATCTTGGGTGGGCCACATTTTTC 4160
1116  rGlnGlnLysAspValLysThrThrGlnProIleGln.....SerValG 1131
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4161  AGAACATATAAAGACTG...TCTAATCCCTCAGGAGAACAGACTTTC 4207
1131  LylsProSerSerIlelleLysAsnValThrThrLysProSerAlaThr 1147
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4208  GATCACTT.....TTAGAAATTTAGAAACTCAGGCCACTTGGTACA 4248
1148  AlaLysTyrThrGlnLysGlnSerGlnProGlnLys..... 1160
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4249  GAAATGATACTGGATTTTCTTCTGAGGTTAAAGAGATTGGAATGAGCC 4298
1161  ....LeuGlnLysleuProLysGlnAlaSerHisGlnleuMetGlnHisG 1176
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4299  GTTCTTAAATCAGCTGGAACACAGATCCATCTGTAGACATGAAGAACAAAT 4348
1176  luleuArgSerSerLysLysSerAlaSerSerGlnLysGly..... 1189
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4349  CGACACAGATCTCT...GGACACAGCAAGTCTGTAGTTATTCACCAAGTCA 4395
1189  ..... 1189
4396  GTGGAAGAGGAGGATGCTTCAATTCAGAGCATCTTCACTGCTGCT 4445
1190  .....ArgAlaLysAspArgGlnHisSerGlySerGln... 1200
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4446  TGATGCTGTACCCAGAACACCTCGAGAGAAAGATGTTCTGCATCTT 4495
1201  .....LysAsp.....AsnProAspLysArg...Lys 1208
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4496  CTCCTGAAGATGAAGATGTTTACCCAGAACTCCATCAAGAGAGAGCG 4545
1209  SerGlyAlaGlnPro.....AspLysGlnSerThrValAspArgLys 1222
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4546  TCTGGGTCTTCTCCAGAGACTTAGAGATGGTCTGGACTCCCTCGAGGCA 4595
1222  uSerGlnGlnGlnHis.....P 1228

```

```

4596  CAGCCTGTCTGGTCTCTCTCGAATGAAGATATACCTAGAACGCCAT 4645
1228  helysThrleuSerGln.....SerSerlysgluthrThrThrSerGln 1242
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4646  TTAGAGGAGAGAGCGAATGTGATCTTCCCGAGAACGAAAGCTTTCCT 4695
1243  LysHisGlnSerValArgGlySerSer.....As 1252
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4696  CAGACTCTCTAGGCGGAGAGATGCTTCCATCATCCAGAGCTCAACAA 4745
1252  nLysAspPheThrProGlnArgAspLysLysValAspTyrAspSerArg 1269
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4746  CAATGCTTTCACCCCGAGAGAAAGAGGGGTGCAATCATCACTTG 4795
1269  spTyr.....Ser 1271
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4796  ATCAGAAAAGTGTGGCTCGAGCTCCCTGGGGCAGAGAACTGTTCCGGA 4845
1272  SerSerLysArgArgAsp.....GluArgG 1280
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4846  TCTCTCAAGAACTTGATGTGAACCCAGTGCATCCCTCAGAAAGAG 4895
1280  yGlnleu.....AlaArgArgLysAspSerProProArgG 1292
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4896  TGACTCAAGACTTCTCCAGATTCTAAAGCCAGAACAGAACCCCACTTC 4945
1292  LylsGlnSerLeuSerGly.....GlnLysSerLys 1302
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4946  GGCAGAGAGATCGGTCTGATCATCTCCAGAGTTGACAGCAATCTCGA 4995
1303  leuArg.....GlnGluArgAspLe 1309
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
4996  CTATCCCTCGGCCAGTAGGCTGTGTTCTCCCTCGAAGTGAAGAAATAA 5045
1309  uProLysLysGlnAlaGlnSerLysSerAsnSerSerProProArg 1326
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
5046  GCCAAGACAGCACCCAGGCGACAGATGTTGTGATTCCTCTCTGAAAC 5095
1326  sPLysLysProHisAspHisLysAlaProTyrGlnLysArgProCys 1342
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
5096  CTAAAGCTCCAGCCCCCTCGGCCCTTCCAGACAGACAGATCAAGTTCA 5145
1343  GlnGlu.....ThrLysProValAspLysAsnSerGlyLysG 1355
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
5146  TCAAGCAAGAGCAGAGGCCCTTCTCTGAAGAGAGCAGACTACCGAGTC 5195
1355  uArgGlnLysHisAlaAlaGlnAlaArgAsnGlyLysGlnSerSerGly 1372
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
5196  CTCCTGTGAACATCCGCCCAATCCAAATCGTCGAGAGATTCAGAGT 5245
1372  laAsnGlyHisValLysLeuThrArgGlnThrLeuProTyrArgArgSer 1388
    |||:|||||: |||:|||||: |||:|||||: |||:|||||:
5246  CATCACAGAGACCCAGAGACCAAGTCTGTACACCACTCGACGTGCGAGC 5295
seq_name: /SIDSI/gcgsdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA107156
seq_documentation_block:
ID      AA107156 standard; DNA; 24387 BP.
XX
AC      AA107156;
XX
DT      21-NOV-2001 (first entry)
XX
DE      Human reproductive system related antigen DNA SFG ID NO: 9844.
XX
KW      Human; reproductive system related antigen; reproductive system disorder;
XX      cancer; gene therapy; ds.
XX
OS      Homo sapiens.
XX
PN      WO20015320-A2.
XX

```



```

13971 .....GAATGAAGAAATCTCGCATTTCTTCAAGCGGAGCGAGTCTC 13929
      ||| ..... ||| ..... |||
623 rPtyrGluSerTyrTyrLysGlyTyrAlaValAlaGlnProArgPro 639
      ||| ..... ||| ..... |||
13928 TC.....TCTTCACCGAGGTCC 13912
640 SerAlaAsn.....ArgGluAspPheSerProGluArgLe 651
      ||| ..... ||| ..... |||
13911 AAGGCAAAATCTCGCTTCTTCTTGAAGCGGACGCTTTCAGGGTCTTCCC 13862
651 uLeuProLeuAsnIleArgAsnSerProPheThrArg.....GlyA 665
      ||| ..... ||| ..... |||
13861 ATGCCCTTAAGCAAAAGTCACAGACACCCAGCGGCGAGTCGCTGGAT 13812
665 rGArgGluAspTyrAlaAlaGlnSerHisArgAsnArgAsnLeuGly 681
      ||| ..... ||| ..... |||
13811 COTCCCAAACTAAAGCTTAATCTAGACGCCACCGACGCGAGTCGCTCC 13762
682 GlyAsnTyrPro.....GluLysLeuSerThrArgAspSe 693
      ||| ..... ||| ..... |||
13761 AGTCTCTCCGCGCACAATAACGAAATCTAAGACACCATCAAGCAAG 13712
693 HisAsnAlaLysAsp...AsnProLysSerLysGluLysGluSerGluA 709
      ||| ..... ||| ..... |||
13711 TCATTCGAGTTCATCTCTCTCATCTTAAGTGA.....TCGGGA 13671
709 snValProGluLysAspLysGlyAsnLysHisLysHisArgLysArg 725
      ||| ..... ||| ..... |||
13670 CACCAACGCGAGGAGGTCATTAACAGTCCCGAGGCCAATGAGCAATCT 13621
726 ArgAsnGluGluGlyGlyGluSerGluSerPheLeuAsnProGluLe 742
      ||| ..... ||| ..... |||
13620 GTAAAGCCACAGAGAGGAGCTGTTTGAATCATCACTGACCTGAGT 13571
742 u.....LeuGluThrSerArgLysSerArgGlySerGlyIleA 756
      ||| ..... ||| ..... |||
13570 GAAATCTAGGACCCCTTCTAGACATAGCTGCTCAGGGTCTCTCTC... 13525
756 spGluThrLysThrAspThrLeuPheValLeuProSerArgAspAspAla 772
      ||| ..... ||| ..... |||
13524 .....CCTAAGCTGAATCTAGC 13507
773 ThrProValArgAspGluProMetAspAlaGluSer..... 784
      ||| ..... ||| ..... |||
13506 ACACCTCCAGACAGAGCCCATCTAGTCTATCATCCACAACCAAACT 13457
785 .....IleThrPheLysSerValSerAspLysAspLysArgGluL 798
      ||| ..... ||| ..... |||
13456 GAAGGCAATTAATATCACCAAGCAAGAACAGCCATCTGCTCTCTCTC 13407
798 ysAspLysProLysValLysSerAspLysThrLysArgLysSer..... 812
      ||| ..... ||| ..... |||
13406 CAAGTCTAGTAGGGTGAAGCTGAGCAACAACTCACGCGGAAGACAGATCA 13357
813 .....AspGlu 814
      ||| ..... ||| ..... |||
13356 GTATCTCCCTGCTCCAAATGTGAATCCAGATTTGCCAAGATACAGTCA 13307
814 ySerIleThrAlaLysLysAspAsnValLeuLysProSerLysGlyProG 831
      ||| ..... ||| ..... |||
13306 TTCTGGGCTCTCTCTCAACAGATACCAAAAGTGAACCTGAAACACCGCCA 13257
831 InGluLysValAspGluLysAspArgGlyLysSerProArg..... 843
      ||| ..... ||| ..... |||
13256 GACAAGATCACTCAGGGTCTATTTCACCATACCCCAAAAGTAAAGGCCCA 13207
844 .....SerGluProProLeuLysLysAl 851
      ||| ..... ||| ..... |||
13206 ACTCCACGCGGCGCAAGTCTTCTGTGATCAAGATCACCATGTCCTCCAGA 13157
851 aLysGluGluAlaThrLysIleAspSerValLysProSer..... 864
      ||| ..... ||| ..... |||

```

```

13156 GAAGTCTAAAGACTACTACTAGTTCAAAGTTGCCCTGATCCCTCTCTCT 13107
865 .....SerSerSerGlnLysAspGluLys.....Val 873
      ||| ..... ||| ..... |||
13106 GTGCAGAGACTAAATCTACACACCCAGCGGAGAGCTATTGTTGGTGC 13057
874 ThrGlyThrProArgLysAlaHisSerLysSerAlaLysAspThrArgAr 890
      ||| ..... ||| ..... |||
13056 TCATCTCTGCAACTGAAGAGACATCTCAAACTTCCACGACACAGATC 13007
890 gGln...SerGlnProArgThrArgArgSerLysArgThrValProLysP 906
      ||| ..... ||| ..... |||
13006 TCATCTTCAAGTCCAGAAAGTGACAGAGTCAATTCAGATACCACTTC 12957
906 hSerSerGlnLysSerGlnProValArgThrArgArgProArgSerLeu 922
      ||| ..... ||| ..... |||
12956 TGCAGAGCAAAATCTCAACATCACCTAAGGAGGTGGTGCAGGTCTTCA 12907
923 ArgLysIleAsnTyrIleuIleAlaArgGluLysAsnGluArgGluLysAr 939
      ||| ..... ||| ..... |||
12906 TCTCCAGTCACTGAGCTGGCATCCAGATCTCCAAATTAAGCAAGATGAG 12857
939 gLysLysSerVal.....AspL 945
      ||| ..... ||| ..... |||
12856 TGAATTCACGCGAGTCTATGTTGAATCTGGAATGCTCTCCGAGCACA 12807
945 ysAspPheGluSerSerSerMetLysIleSerLysValGluGlyThrGlu 961
      ||| ..... ||| ..... |||
12806 GCAGGTTCAGTCACTGACTCTCTCATATCTTACAGTGAAGTCAAGATTC 12757
962 IleValLysProSerProLysArgLysMetGluLysAspValGluLysLe 978
      ||| ..... ||| ..... |||
12756 CTTCTGGGCGCAGAGTAAATGAGACTGCTGAATCAAA...GAGAAAT 12710
978 uGluArgThrProGluLysAspLysIleAlaSerSerThrThrProAla 995
      ||| ..... ||| ..... |||
12709 GGCTTACCCCTCAGAGAGAT.....GCTACTGATCACCTCCCTAAGAC 12666
995 yslLysIleLysLeuAsn..... 1000
      ||| ..... ||| ..... |||
12665 AGAAAGCAAAATTAAGTCTTCCAGTACAGGATAGGCTGAGTCTTCA 12616
1001 .....ArgGluThrGlyLysIleGluLysAl 1010
      ||| ..... ||| ..... |||
12615 CTGTGATTTCAAAGACACTTGAACCCCGCCAGGAAGAAAGTGTGTC 12566
1010 aGluAsnAlaSerThrThrLysGluProSerGluLysLeuGluSerThrs 1027
      ||| ..... ||| ..... |||
12565 TGGGTCTATCTCCAGAAACAAAGAGCAAAATAGTGCATTGCTTCA 12516
1027 eSerLysIleLysGlnGluLysValLysGlyLysAlaLysArgLysVal 1043
      ||| ..... ||| ..... |||
12515 GCCAAGATGAAGGTTAAATGAGGTGTAGAGAAAGCTGAAGAACCCGCA 12466
1044 AlaGlySerGluGlySerSerSerThrLeuValAspTyrThrSerThse 1060
      ||| ..... ||| ..... |||
12465 GGCCAAATCCCTGCTCATTTGCTTCAGAACTTAAGAAATGTCCACAAG 12416
1060 rSerThrGlyLysSerProValArgLysSerGluGluLysThrAspThrL 1077
      ||| ..... ||| ..... |||
12415 TTACTTTGAATCATCTCTC.....GAAGTGAAGAAA 12384
1077 ysArgThrValIleLysThrMetGluGluTyrAsnAsnAspAsnThrAla 1093
      ||| ..... ||| ..... |||
12383 GGCCTCTGTGCTTGTGACTCTGATCAGAGCAGTACACAGGCTTCTTGG 12334
1094 ProAlaGluAspValIleIleMetIleGlnValProGln.....SerL 1108
      ||| ..... ||| ..... |||
12333 GAAGCA.....GTAAGATCCCTTCATAGGCTTCATC 12302
1108 sTrpAspLysAspAspPheGluSerGluGluGluAspValLysThrThrG 1125
      ||| ..... ||| ..... |||
12301 TTGGGGTGGCCCATTTTCTCCAGAACATTAAGAACTG...TCTAACT 12255

```

THIS PAGE BLANK (USPTO)


```

111 TTCAATGAAATCTCGAAACTGAAAGTGCATGAAATAGTGAACCATCAGC 160
967 rOLysArgLysMetGluGluGluValGluLysLeuGluArgThrProGlu 983
161 CAAAGCGGAAATGGAACTGATCTGATGAAAAATGGATAGACCCCTGAA 210
984 LysAspLysIleAlaSerSerThrThrProAlaLysLysIleLysLeuAs 1000
211 AAGCAAAATTT...TCTTAAAGTGGCCAGCCCAAAATCAAACTGCA 257
1000 nArgGluThrGlyLysLysIleGlyAsnAlaGluAsnAlaSerThrThrL 1017
258 CAGAGAAACTGGGAGAAATTTGGAAATGACAGAAATATATCAAAACACA 307
1017 ySgIProSerGluLysLeuGluSerThrSerSerLysIleLysGluGlu 1033
308 AAGAACCTCTGAAAAATTTGGAGTCACATCTGACAAAGTTAAACAGAA 357
1034 LysValLysGlyLysAlaLysArgLysValAlaGlySerGluGlySerSe 1050
358 AAAGTCAAGGAAAGTCAGACGAAAGAGTGAAGTGAAGTGAAGTCAAG 407
1050 rSerThrLeuValAspTyrThrSerThrSerSerThrGlyLysSerPro 1067
408 CTCACCTCTGTGTGATACACCGATGACGATCACTGAGGACGTCTGTG 457
1067 aLArgLysSerGluGluLysThrAspThrLysArgThrValIleLysThr 1083
458 TCGGAAATCTGCAAGAAAAACAGATACAAAGCGAAGCTGTGTTAAAAAC 507
1084 MetGluGluTyrAsnAsnAspAsnThrAlaProAlaGluAspValIleL 1100
508 ATGGAAGATATATATATATGACAAATACCGCGCACTGAAGATGTTATCAT 557
1100 eMetIleGluValProGluSerLysTyrAspLysAspSerProGluSerG 1117
558 TATGATTTGAGGTTCTCAATTCCTCAATTCCTCAATTCCTCAATTCCTG 607
1117 LuGluGluAspValLysThrThrGluProIleGluSerValGlyLysPro 1133
608 AAGAAGAAAGATGTTAAATTCACACAGCCTATATCAAGTGTAGAAAACT 657
1134 SerSerIleIleLysAsnValThrThrLysProSerAlaThrAlaLysTy 1150
658 GCTAGTGTATATATATATATGTTAGTCAAAAGCCATCAAAATATATGCA 707
1150 rThrGluLysGluSerGluGluProGluLysLeuGluLysLeuProLysG 1167
708 TCCTGAGAAAGAAAGTGAAGCCATCCGAGAAATTCAGAAATTCACCAAG 757
1167 LuAlaSerHisGluLeuMetGluHisGluLeuArgSerSerLysGlySer 1183
758 ACGTGAAGCATGAATTCATACACATGAGTTAAAGTTCAAAAAACTCT 807
1184 AlaSerSerGluLysGlyArgAlaLysAspArgGluHisSerGlySerG 1200
808 GCATCTAGTAAAAAGGAAAAAGCAAGATGAGATATATTCAGTGTGGA 857
1200 uLysAspAsnProAspLysArgLysSerGlyAlaGluProAspLysGlu 1217
858 AAAGGAGAACCTGAAAAAGAGAAAGACAGCATCCAGCCAGAGAAAGAGA 907
1217 eThrValAspArgLeuSerGluGluGluHisAspLeuThrLeuSerGlu 1233
908 GTAAATTTGACCGTCTGAATGACAGGAAATTTAAAGTCTGCTCA 957
1234 SerSerLysGluThrArgThrSerGluLysHisGluSerValArgGlySe 1250
958 TCTTCCAAAGAGGCTAGAGACGATGAATTAACATGATTCACACGCGCTTC 1007
1250 rSerAsnLysAspPheThrProGlyArgAspLysLysValAspTyrAsp 1267
1008 CTCAAATTAAGACTTCACTCCCAATAGAGACAAAAAACTGACTATGACA 1057

```

```

1267 eArgAspTyrSerSerSerLysArgArgAspGluArgGlyLeuLeuAla 1283
1058 CCAGAGAGATATTCAGATTCCTCAAAACCTAGAGATGAAAGATATTAACA 1107
1284 ArgArgLysAspSerProProArgGlyLysGluSerLeuSerGlyGlu 1300
1108 AGACGAAAAAGACTCTCTCTCGAATTAAGATTTCTGCATCTGCAGACAGA 1157
1300 sSerLysLeuArgGluGluArgAspLeuProLysLysGly...AlaGlu 1316
1158 AATTAACCAAGGAGAGAGAGATTTGCTTAAAGGAGGAGGAGATT 1207
1316 eLysLysSerAsnSerSerProProArgAspLysLysProHisAspHis 1332
1208 CCAAAAAAGTATTCCTAGTCCCTCAAGAGACAGAAAAACCTATATATCAG 1257
1333 LysAlaProTyrGluThrLysArgProCysGluGluThrLysProValAs 1349
1258 AAAGCCACTTATGATATTAAGCGCCAAATGAAAGACAAATCTGTAGA 1307
1349 pLysAsnSerGlyLysGluArgGluLysHisAlaGluAlaArgAsnG 1366
1308 TAAAAATCTTGTAAAGATCGTGAAGACATGTATAGACGAAAGACA 1357
1366 LysGluSerSerGlyAlaAsnGlyHisValTyrLeuThrArgGluThr 1382
1358 ATAAAGAGTCAAGTGCATTAATTAATCTTATATATCTTATACCCACAGAG 1407
1382 rLeuProTyrArgArgSerTyrLeuGluGlyArgTyrPargAlaProS 1399
1408 ACACAGGTTGAAAAAGACCAATATCTGGCAAAATTGACAAAGATCTGT 1457
1399 eSerArgAsnProSer 1404
1458 CAAGCTTAACCCAGT 1474

seq_name: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: PCT-US93-11310-11
; seq documentation block:
; Sequence 11, Application PC/TUS9311310
; GENERAL INFORMATION:
; APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11310
; FILING DATE: 19-NOV-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: FP-CJ 9790
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9901
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1800 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
;
PCT-US93-11310-11

```

```
alignment_scores:
  Quality: 1838.50
  Ratio: 4.113
  Gaps: 5
Percent Similarity: 91.224
Percent Identity: 75.510
```

```
alignment_block:
US-09-811-045A-1 x PCT-US93-11310-11
```

Align seg 1/1 to: PCT-US93-11310-11 from: 1 to: 1800

[illegible][illegible]

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Hovey, Williams, Timmons & Collins
;; STREET: 2405 Grand Boulevard, Suite 400
;; CITY: Kansas City
;; STATE: Missouri
;; COUNTRY: U.S.A.
;; ZIP: 64108
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/931,999
;; FILING DATE:
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/710,561
;; FILING DATE: 19-SEP-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Collins, John M.
;; REGISTRATION NUMBER: 26,262
;; REFERENCE/DOCKET NUMBER: 25043-A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 816/474-9050
;; TELEFAX: 816/474-9057
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 6755 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHEICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Staphylococcus aureus
;; STRAIN: UT0007
;; US-08-931-999-4

alignment_scores:
Quality: 477.00 Length: 995
Ratio: 0.958 Gaps: 40
Percent Similarity: 50.050 Percent Identity: 21.508

alignment_block:

US-09-811-045A-1 x US-08-931-999-4 ..

Align seg 1/1 to: US-08-931-999-4 from: 1 to: 6755

```
550  ArgPrtYrArgSerArgSerArgSerProProGluPheArgGlyInse 566
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2890  CGAGCGCAGCAAAAGAAACAGCA..... 2913
566  rProThrLysArgAsnValProArgGluGluLysGluArgGluTyrPhea 583
      ::||| ||||| ||||| ||||| ||||| ||||| |||||
2914  .... AGAGAAAC... CCGAGGAGAAAGCAAAAGCAGACAAAGCAGCA 2953
583  snArgTyrArgGluValProProProTyrAspIleLysAlaTyrTyrGly 599
      ::|||
2954  GGAGA..... 2958
600  ArgSerValAspPheArgAspProPheGluLysGluArgTyrArgGluTr 616
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
2959  ..... GAGAAAAAGAGCAGACAGACAC 2981
616  pGluArgLysTyrArgGluTyrTyrGluLysTyrTyrLysGlyTyrAlav 633
      ::||| ||||| ||||| ||||| ||||| ||||| |||||
2982  CAGAAAGCCCGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAG 3025
633  alGlyAlaGlnProArgProSerAlaAsnArgGluAspPheSerPro... 648
      ::||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
3026  CAGGGCGCAGCGGACAGGAAAAACAAAGAAAAACAGCCGACCCAGG 3075
649  ..... GLuArgLeuLeuProLeuAs 655
3076  CAGCAAAAGGGGGAGAGAAAAACAAAAACAGAAAAACAGACACAGCAC 3125
655  nIleArgAsnSerProPheThrArgGlyArg..... 666
3126  CACAAGAAAGCAACCCAAAAAGAAAGAAAGGAGGAAACACAAAGAA 3175
667  ..... GLuAspTyr 669
3176  CGAAAAACAAACGGCCCAACAGCAACCAAAACAAAGCCGAGAAAGACGA 3225
670  AlaIaGlyInSerHisArgAsnArgAsnLeuGly..... 681
3226  GCGAAAGGGGCAAGCAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3275
682  ..... GlyAsnTyrPro.GluLysLeuSerThrArgAspSer 693
3276  AACCCAAAAAGAGAGAAAGCAACCCAGCAAGGCGCAGCAGCAAGAGCA 3325
694  HisAsnAlaLys... AspAsnProLysSerLysGluLysGluAs 709
3326  CACGAGAAAAAGCACACAAACCAAGCCAGCAGCAGCAAGCAAGCAAG 3375
709  nValProGlyAspGlyLysGlyAsnLysHisLysLysHisArgLysArg 726
3376  GAGGACAGAGAAAGAAAGGAGGCGCACAAGAAAGAAAAAGAAAGAA 3425
726  rGAsnGluGluLys..... GlyLysGluSerGluSerPheLeuAsn 739
3426  ACAAAAGCAAAAAAGAGGAGGAGAGAAAGAAAGAAAGCAACAAAGAA 3475
740  ProGluLeuLeuGluThrSerArgLysCysArgGlySerGlyIleAs 756
3476  ACCGAGCAGAAAGAAAAAGGAAAAAGAAAAAGAAAAAGAAAAAGAA 3525
756  pGluThrLysThrAspThrLeuPheValLeuProSerArgAspAspAla 773
3526  GAAGCAAGACGAGCA..... 3541
773  hrProValArgAspGluProMetAspAlaGluSerIleThrPheLysSer 789
3542  .. GAAAAAAGAAAGAAACAGCAGCAAGCAAGCAAGCAAGCAAGCAAG 3589
790  ValSerAspLysAspLysArgGluLysAspLysPro..... 801
3590  GAAAAAACGAGAACAGAGAGAAAAAGAAAAAGCAACAGCAACCCGAG 3639
802  ..... LysValLysSerAspLysThrLysArgLys 812
3640  CGACCAAAAAAAGAAAGCCGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3689
812  erAspGlySerAlaThrAlaLysLysAspAsnValLeuLysProSerLys 828
3690  AAGAAAAAAGAAAGAGGAGCAAA..... GCGCAGAAAAAAC 3727
829  GlyProGluGluLysValAspGlyAspArgGluLysSerProArgSerG 845
3728  GGAGGAAACAGAAAGAAAGCAAGAAAAAGCAAGAAAAAGCGGGA 3777
845  uProProLeuLysLysAlaLysGluGluAlaThrLysIleAspSerValL 862
3778  CAGACCAAAAAAAGAAAGGAGCAGGAGAAAGCAAGAAAGCAAGCAAG 3827
862  ysProSerSerSer..... SerGlnLysAspGluLysValThrGly 875
3828  AAAAGAGAGAGCAGCAAGAAAGAAAGAAAGAGAGCAAGCAAGAAAG 3877
876  ThrProArg..... LysAlaHisLe 882
3878  ACAAGAGAGAGAGAGAGAGAGAGCAAGCAAGCAAGCAAGCAAGAGAC 3927
```



```

347  rCysTyrArSerIleAsnArgGlyArgHisHisSerGluArgSer..Gln 363
    ||||| ||||| :::: ||||| ||||| :::: |||||
66882  CTGCCCCAGATCGCGCTCCAGTCGGCAGATCTTCTGTCAAACGGCTCAG 66833
364  ArgThrGlnSerProSerIleuProAlaThrProCysPheValProValPr 380
    ||||| ||||| ||||| ||||| :::: |||||
66832  CGAACA.....CCGAAAAACCGAGTAGAGCCCCGGCCCC 66798
360  oProProProLeuTyrProProProProHisThrIleuProIleuPro... 395
    ||||| ||||| ||||| ||||| :::: |||||
66797  ATCCCAACCACTGTCCGCCCTGACCGCGGACACGAACCCACCCCGCC 66748
396  ....ProGlyValProProGlnPheSerProGlnPheProSerSer 410
    ||||| ||||| ||||| ||||| :::: |||||
66747  AGAGACCGGGGATCCCGACACA.....CCCAAGAGCCTCGCGAC 66707
411  GlnProPro.....ThrAlaGlyTyrSerValProProPr 422
    ||||| ||||| ||||| ||||| :::: |||||
66706  CGGAGCCGAGCGCTTCCAACCGGACACAGTTTCGGCGTCCGCTGCC 66657
422  oGlyPheProProAlaProAlaAsnIleSerThrAlaCysPheSerProG 439
    ||||| ||||| ||||| ||||| :::: |||||
66656  AGCACCACCGCGCTCTCCACAG.....CGGCGCG 66625
439  ly...ValProThr.....Ala 443
    ||||| ||||| ||||| ||||| :::: |||||
66654  GGAGAGCCCAAGCAAAACCGCGCATCCAGCTGGAAATATCCGACGCT 66575
444  HisSerAsn.....ThrMetProThrThrGln..... 452
    ||||| ||||| ||||| ||||| :::: |||||
66574  CACCCAAATAGCAACAATAGCGCTGACCGCTGAGCGCTGAGCGCTGCCGT 66525
453  ...AlaProLeuLeuSerArgGluGluPheTyrArgGluGlnAsnAspL 468
    ||||| ||||| ||||| ||||| :::: |||||
66524  GTTCTCCCCGACACGATCAACGGGACCGCTGGAGATCCACGCTGATTC 66475
468  ysgLyArgGluSerLysPheProTyrSerGlySerSerTyrSerArg... 483
    ||||| ||||| ||||| ||||| :::: |||||
66474  CGAATCGGAGCGCTGTGTCGCACTCTGCTGCTGTTTCGAGGCGTT 66425
484  .....SerSerTy 486
    ||||| ||||| ||||| ||||| :::: |||||
66424  GTTCGAGGATGACGTGCTGTGTCGCGCTGATCCCGACAGCAGCAGAC 66375
486  rThrAspSerSerGlnGlyLeuAlaGlnHisIleHisIleAlaLeuThrL 503
    ||||| ||||| ||||| ||||| :::: |||||
66374  CCCACCCCGCGAGATGGCTGTGCCG.....GCCAGGGGGTGT 66334
503  ePrProSerAlaAlaHisThrIleuAspLeuLeuHisAsp..... 515
    ||||| ||||| ||||| ||||| :::: |||||
66333  CTCCCTGAGGAGCTGGA.....CATCCCCCGCGGACCAAT 66299
516  .....HisProHisProProGlnGluAlaGluAlaIar 526
    ||||| ||||| ||||| ||||| :::: |||||
66298  CCACCTCCGACCTAGGTTTCATCCATGCACTGCGGGCAGCTGCCG 66249
536  gSerAlaMetIleVal.....HisMetProAspL 536
    ||||| ||||| ||||| ||||| :::: |||||
66248  TGCCTGATGCGCATCATCATTGTATCAACAGGAGACCGCGCGCGCC 66199
536  euMetAspIleAlaHisIleAspSerArgSerPro..... 547
    ||||| ||||| ||||| ||||| :::: |||||
66198  CTGCGTATGACCAATATTCGACTCAGCAGCCCAACCAACAGGCGCAT 66149
548  .....ProTyrArgArgTyrArgSer..... 554
    ||||| ||||| ||||| ||||| :::: |||||
66148  CGCTATCCCGCCCTGCGCGTAAAGTGGAATGAGAGCCTGCGCCTCAT 66099
555  .....ArgSerArgSerProProGlnPheArgGlyG 565
    ||||| ||||| ||||| ||||| :::: |||||
66098  GGATCACCAGCAAGCTGTGTTCCATCCCATCGGCTCCAGCGGCTCCAC 66049

```

```

565  lnsErProThrIysArgAsnValPro..... 573
    ||||| ||||| ||||| ||||| :::: |||||
66048  AGACACCGCAAAACCAAGCAGCTGCCACAGCCTGCGGATCACCCTGCT 65999
574  ...ArgGluGln..LysGluArgGluTyrPheAsnArg.....T 585
    ||||| ||||| ||||| ||||| :::: |||||
65998  GCGAGGCGCCCATTCGGCGCGCTCAAAACATTCGACGACCAACATCTGAT 65949
585  yTrArgGluValProProProTyrAspIleLysAlaTyrTyrGlyArgSer 601
    ||||| ||||| ||||| ||||| :::: |||||
65948  ACCGGGCTAAC.....CCGACAGACCGGCAACCGCGTCCCGCTGGC 65905
602  ValAspPheArgAspProPheGluLysGluArgTyrArgGluTyrPolu 618
    ||||| ||||| ||||| ||||| :::: |||||
65904  CCGGGCATCCGACAGCGCTTCCACACACCAAGCCCGCGCC..... 65863
618  gLysTyrArgGluTyrPyrGluLysTyrTyrLysGlyTyrAlaValGlyA 635
    ||||| ||||| ||||| ||||| :::: |||||
65863  ..... 65863
635  lAglnProArgProSerAlaAsnArgGluAspPheSerPro..... 648
    ||||| ||||| ||||| ||||| :::: |||||
65862  .....TTCCTCCCAACCGGTGCC 65845
649  .....GluArgLeuLeuProLeuAsnIleArgAsnSerPr 660
    ||||| ||||| ||||| ||||| :::: |||||
65844  ATCCGACGCCCGCCAAAGACTTGCACCGCCCATCCGCCCAAGCCCC 65795
660  oPheThrArgGlyArgArgGluAspTyrAlaAlaGlyGlnSerHisArgA 677
    ||||| ||||| ||||| ||||| :::: |||||
65794  GTTACGCGCAAAACATCCAGCAACATCCCGCGGTGCCAT...CACGCTC 65748
677  snArgAsnLeuGlyLysAsnTyrProGluLysLeuSerThrArgAspSer 693
    ||||| ||||| ||||| ||||| :::: |||||
65747  AC.....ACCAACCGCAAGAC 65731
694  HisAsnAlaLysAspAsnProLysSerLysGluLysGluSerGluAsnA 710
    ||||| ||||| ||||| ||||| :::: |||||
65730  CAGATCACACTCACAGACCGAGCTGCTTGACCCGCGCAGGTGAG..... 65686
710  lProGlyAspGlyLysGlyAsnLysHisLysHisArgLysArgArg. 726
    ||||| ||||| ||||| ||||| :::: |||||
65685  .....GCCACCAAGCAGACGACGACGCGCTA 65658
727  ....AsnGluGluLysGlyGluGluSerGluSerPheLeuAsnProGlu 741
    ||||| ||||| ||||| ||||| :::: |||||
65657  TTCACCGTCACCGCAGAGACCTCGAAACCAACGAAATAGCAACCCGAC 65608
742  .....LeuLeuGluThrPse 746
    ||||| ||||| ||||| ||||| :::: |||||
65607  CGAAAAACAGCGCTCCCGCGCTGCCATTACCTAGTAACCTCGAAACCT 65558
746  rArgLysCysArgGlySerSerGlyIleAspGluThrLysThrAspThrL 763
    ||||| ||||| ||||| ||||| :::: |||||
65557  CCGGGCGCTCTGT.....GATGAACCGCGCGCGCTGTGTG 65523
763  euPhe..ValLeuProSerArgAspAspAlaThrProValArgAspLPr 779
    ||||| ||||| ||||| ||||| :::: |||||
65522  TGTGTCATGAGCCCGCGCAAC.....ACGCGGTCCGCTCCCGC 65482
779  oMetAspAlaGluSerIleThrPheLys..... 788
    ||||| ||||| ||||| ||||| :::: |||||
65481  CAACGAAACGGATCAATCCCGCGCTCCACAGGTTTCCCAAGCAGCTT 65432
789  .....SerValSerAspLysAspLysArgGlu 797
    ||||| ||||| ||||| ||||| :::: |||||
65431  CAGACGCAACCGCTCTCGGATCCATGCGCAACCGCTCAGCGGGGGA 65382
798  ...LysAspLysProLysValLysSerAspLysThrLysArgLys..... 811
    ||||| ||||| ||||| ||||| :::: |||||
65381  ATCCGAAAGAAACCGCATCAATCTGCGGATCTCGCAGGAAGCAGCC 65332
811  ..... 811

```

```
65331 CTCACACATAGAGGTCCAGACGATCGGATCCGATCGAACAAC 65282
812 .....SeraspGlySerAlaThrAlaLysLys 820
65281 CCGCCTGATCCACGACGATGTCGGGAAGTCCGACCGCATCCAAAG 65232
821 AspAsnValLeuLysProSer..... 827
65231 CCCCCGGCACCAAAAGCTCTCCGGCGAGACACCCCAACCCGG 65182
828 ...Lys.GlyProGlnGluValAspGlyAspArgGluLysSerPro 842
65181 ATAAAGACAGCCCATCCACGATCACAACGGATCGTCAGACGATCAGC 65132
843 ArgSerGluProProLeuLys..... 849
65131 CCGCCACAGCCGCGGTGCGGAGTCCGAGCTCCACGCCCAATCAGT 65082
850 ...LysAlaLysGluGluAlaThrLysIleAspSerValLysProSer 865
65081 TTGTGAGCAAGAACGACCAACG....CCGCGGTGTGCGATATGC 65038
865 erSerSerGlnLysAspGluLysValThrGlyThrProArgLysAlaHis 881
65037 GAACACGCGCATAGCCGCAACCGACGCCCAACCCCGCAAGCCGT 64988
882 SerLysSerAlaLysAspThrArgArgLysSer..... 892
64987 TCCGAACCTCCACACCGACACAGATATAAACCMAATCCCGGAACCA 64938
893 .....GlnProArgThrArgArgSerLysArgThrValProLys 905
64937 CGATCGGCCCGACCTGACCGCAACCGCAATGCCACGACCGCGAAGC 64888
906 .....Thr 906
64887 ATGCTCGGCGCAACCAACGACGCTGCTTCATCGAGGTCCGCGCA 64838
907 SerSerGlnLysSerGlnProValArg.....ThrArgArgPro 919
64837 GCTTCGCGACGCGCAACGACGACGCTTGACGTCGCCGACTCGTTGACC 64788
919 oArgSerLeuArgLysIleAsn.TyrLeuIleAlaArgGluLysAsnGlu 935
64787 GCGCTCGCGCCGACGCGGTGCGACGACGCGCAACGACGACGAC 64738
936 ArgGluLysArgLys.....LysSerValAspLysAspPheGlu 949
64737 ACCGCGCTCGCGCATCAAGCGCGAGGCGTTCAGTCAAGTGGACG 64688
949 rSerSerMetLysIleSerLysValGluGlyThrGluIleValLysPro 966
64687 GAACCAAGAGGATTCGCCCGGAACGCAACGCGCATCTCAACACGCGC 64638
966 erProLysArgLysMetGlnGluLysValGluLysLeuGluArgThrPro 982
64637 AACCGCTCGCGCGGACATCGCGGATTTGTT..... 64606
983 GluLysAspLysIleAlaSerSerThrThrProAlaLysLysIleLys 999
64605 .....GCGCGCAATGCGGCTACGCGCGCGCATCGACG 64571
999 uAsnArgGluThrGluLysLysIleGlyAsnAlaGluAsnAlaSerThr 1016
64570 CTCGCGTATCCGCTGCGGTCTCCCAAGACCCCAAGCGAGGACACC 64521
1016 hrLysGluProSerGluLysLeuGluSerThrSerSerLysIleLysGln 1032
64520 GCAAGGACGCC....CGCGTCCGCGGTGCTGACGACGACGCGTCCAG 64477
1033 GluLysValLysGlyLysAla.....LysArgLysValAla 1044
|||||.....:::|
```

```
64476 GAACGCTTCCGCCCGCGTAGTTTCCCTGACCCGCAAGCCCCCGCAACG 64427
1044 agLysGlnGlySer..SerSerThrLeuValAspTyrThrSerThrSe 1060
64426 CGAGAACGAGAGAGAGACAGAAAGACGAGATCCAGATCAAGAGTC 64377
1060 rSerThrGlyLysSerProValArgLysSerGluGluLysThrAspThrL 1077
64376 AACTGTGACAGGTGCCACGACGCGCGCTTCGGGCGGAACACTTTCG 64327
1077 ysArgThrValIleLysThrMetGluIuTyrAsnAsnAspAsnThrAla 1093
64326 CACGTGCTCTGTGTC.....AGCGATCTCAGCAGC 64296
1094 ProAlaGluAspValIleIleMetIleGlnValProGlnSerLysTyrPas 1110
64295 CCGTGGCAATGCCCCCGCCGATGCACACCTCGCGCAGCGGGAATC 64246
1110 pLysAspAspPheGluSerGlnGluAsp.....ValLysThrT 1124
64245 TGCCGAGATGTCGTGACAGACGCGGTCCAACTGGGCCCCGATCCCGCAT 64196
1124 hrGlnProIleGlnSerValGlyLysPro.....SerSerIleLe 1137
64195 CGCAGCTCGCCACTCGACCGCAGCTCCAGGCCACCACTCATCAC 64146
1138 LysAsnValThr.....ThrLysProSerAlaThrAl 1148
64145 AGTCGCGGACCCCGGGGCGCTCCAGCGCGGTGCGCCCGGACACACAG 64096
1148 aLysTyrThrGluLysGluSerGluGln..... 1157
64095 GCGCGCGACCGCGCTCCGCCCAAAATCGCGGCAACCAACCGGCCA 64046
1158 .....ProGluLysLeuGlnLysLeuProLysGluAlaSerHis 1170
64045 GCAGCGCCGATCGCCGGAAATCAACACCTGCCA..... 64011
1171 GluLeuMetGlnHisGluLeuArgSerSerLysGlySerAlaSerSerG 1187
64010 .....TCGCGACGGGCGACTGGGACTCGGCT 63985
1187 uLysGlyArgAlaLysAspArgGluHisSerGlySerGluLysAspAsn 1204
63984 GTCTCGCGCGCAACATCGCGCGGCGAGGACGAC...CGGACTTCAC 63938
1204 roAspLysArgLysSerGlyAlaGlnProAspLys..... 1215
63937 CCGCGCGACGCGCAGCTCGCGGTCTCCGGAAGCAGACGCGCGCGAC 63888
1216 .....GluSerThrValAspArgLeuSerGluGlnGlyHisPheL 1229
63887 GCAGCGCATGATTCGGCGGTTCGTCACATCGAACCAAGATTCGCTC 63838
1229 sThrLeuSerGlnSerSerLysGlu.ThrArgThrSerGluLysHisGlu 1245
63837 CGGCTTCCGACTGCGCGCAACCAACACCCACGACGACGACATGCA 63788
1246 SerValArgGlySerSerAsnLysAspPheThr..... 1256
63787 CCAGCTGGGAACACCTCGCGGGGACCTGTGACGCGCGCACACACGTC 63738
1257 .....ProGlyArgAspLysValAspTyrAspSerArgAspTyr 1271
63737 ACAACCAACGAGCGAGATCCGCGAATCGCGCGCGCGCAACCAATCTG 63688
1271 erSerSer.....LysArgArg 1276
63687 GATCAGGTCACATGCGTTGTGTGCGGACAGTGTCTGGAAACCAAGT 63638
1277 AspGluArgGlyGluLeuAlaArgArgLysAspSerProProArg.... 1291
63637 CCAAGTCTGTGCGAGACGACCGCATG....GCACCACACGACGCTCC 63594
```

```

1292 .....GlyLysGluSerLeuSerLys..... 1298
      |||||
      |||||
63593 GGGGGGGGGTCAACCCAGCCGGTCAAGTGGAGTCCGTCGCAAG 63544
1299 .....GlnLysSerLysLeuArgGluLysArgAspLeuProLys 1311
      |||||
      |||||
63543 TCCCGAGCCGCGGCGTGGCGAGTTCGTCAATTCGCGATCAGTCCGCAAG 63494
1312 LysGlyAlaGluSerLys.....SerAsnSerSerProLys 1325
      |||||
      |||||
63493 AACCCGGCTTGGCAGCGCCGGGAGATTCGTTCCAGTCGACCCGGAAC 63444
1325 gAspLysLysProHisAspHisLysAlaProTyrGluThrLysArgProC 1342
      |||||
      |||||
63443 AGGGAAGTCTGTTCCATCGCAC...GGCCGCTGCCAAGTCTGTTGCGCA 63397
1342 ysgLgLuThrLysProValAspLys.....AsnSer 1352
      |||||
      |||||
63396 AATCTGGCGCAGACCGAGATTCATGAGAGCAGCGCTGGCCAGTGC 63347
1353 GlyLys.GluArgGluLysHisAlaAlaGluAlaArgAsnGlyLysGlu 1369
      |||||
      |||||
63346 TGTCTGCTCGTGAACACCGACCGCATTCGCGCCCGGAAGAGGTGC 63297
1369 eSerGlyLysAsnGlyHisValTyrLeuThrArgGlnThrLeuProTyr 1385
      |||||
      |||||
63296 ACCCGAAGGGAAGTCCGCCGCTGCTTGAAGGAGACACCTTGCACGC 63247
1386 ArgArg 1387
      |||||
63246 GAACCG 63241

```

seq_name: /cgn2_6/prodata/1/ina/6B-COMB.seq:US-09-370-700-1

```

seq_documentation_block:
: Sequence 1, Application US/09370700
: Patent No. 6274350
: GENERAL INFORMATION:
: APPLICANT: Balcz, Richard H
: APPLICANT: Broughton, Mary C
: APPLICANT: Crawford, Kathryn P
: APPLICANT: Madduri, Krishnamurthy
: APPLICANT: Treadway, Patti J
: APPLICANT: Turner, Jan R
: APPLICANT: Waldron, Clive
: TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
: FILE REFERENCE: 50489 DIV1
: CURRENT APPLICATION NUMBER: US/09/370,700
: EARLIER FILING DATE: 1998-03-09
: EARLIER APPLICATION NUMBER: US 09/36987
: NUMBER OF SEQ ID NOS: 39
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 80161
: TYPE: DNA
: ORGANISM: Saccharopolyspora spinosa
: US-09-370-700-1

```

```

alignment_scores:
  Quality: 366.00      Length: 1708
  Ratio: 0.499        Gaps: 83
  Percent Similarity: 42.974  Percent Identity: 19.965

```

Alignment block:

US-09-811-045a-1 x US-09-370-700-1/rev ..

Align seg 1/1 to reverse of: US-09-370-700-1 from: 1 to: 80161

```

23 ProThrIleAspAlaGlu.....AlaTyrAlaIleGlyLysLysGlu 37
|||||
|||
|||
|||

```

```

67729 CCGACCATTCGACCGGACACACCGCGGTGTGCGCCGCCACCGCGCG 67680
37 sProProPhe.....L 41
|||
|||
67679 ACCACAGTTTCGTTCCGTACGGTCTTACGACGATCGGAATCGGCAT 67630
41 euProGluLupProSerSerSerSerGluLupAspAspProIleProAla 57
|||||
|||||
67629 GGCCTGATCAGCCCGACGAAATGACATCGGGAATCAGGCGCGACAGC... 67584
58 GluLeuLeuCysLeuIleCysIlyAspIleMetThrAspAlaValAl 74
|||||
|||||
67583 .....GTCCATCAGGCGCGAGTTCGAGATCGTGGCGAC 67549
74 eProCysCysGlyAsnSerSerCysAspIleCysIleArgThr..... 88
|||||
|||||
67548 CCTTGTCTCGCGGAGCAGCCGCGAGCGGCGGGAAGGAGCGGCGTC 67499
89 .....Thr.LeuLeuLupSerAspLysHisThrCys..... 98
|||
|||
67498 GAACCTGACGACCCCAATCTCGGCGTACCATCACACCTCGCGTGTG 67449
99 .....ProThrCys 101
|||||
67448 AGCTACCGCTACAGTCGACACGACCGGTAGGTGCGCACCGCGGTATG 67399
102 HisGlnAsnAspValSerProAspAlaLeuIleAlaAsnLysPheLeu 118
|||
|||
67398 CACGCT.....CGTCCGATTTGCGGAACTCCTGCA 67367
118 gGlnAlaValAsnAsnPheLysAsnGluThrGlyTyrThrLysArgLeu 135
|||||
|||||
67366 ACATCGGCTCCATCGATACGA.....ATGAAGCGCATCG 67332
135 rGlyGlnLeuProProPheLeuPheLeuValProProArgProLeu 151
|||||
|||||
67331 GAACCCGCGACCG.....CCGACCGAATCCTTG 67300
152 SerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuArgGln 168
|||||
|||||
67299 .....CCCATCCAGCGCCCGCG 67283
168 nAspProValAlaPheArgTyrThrValSerProThrCysSerAspThr 185
|||
|||
67282 CGATGCTC.....AGCCACCTCC 67263
185 yThrAlaGly.SerCysSerAspSerGlyThrLeu...SerArgLeuP 200
|||||
|||||
67262 CGATCGCGGAGAGCAGCACCGCATCGAGGCTTACCGCGCGCATACC 67213
200 oAlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaPro 217
|||||
|||||
67212 CGCACCATCAACCCGATCGGCCCAACGACGCGCAGCTGCTCTCACAG 67163
217 roValSerGlyAsnProSerSerAlaProAlaProValProAspIleThr 233
|||||
|||||
67162 CAGCCACCGCCAC.....ATGCGACCG.....CCAGAGCGCAC 67128
67127 GCCTGCATC..... 67119
250 gAspSerAspAsnLysLeuLeuProAla.....AlaAlaLeuThrS 264
|||||
|||||
67118 .....AACCGGCGACGACCCCGCACCATCCGAGCGCATCCGCGCA 67079
264 eArgLupHisSerLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGlu 280
|||||
|||||
67078 ACGACACACATCCCGCGCAACCGCGCGGCGGACTGCGCGACCGCATG 67029
281 GluLysGlyValProGlyThrSerProTyrAsnSerIlePheValGly 297
|||
|||
67028 CCGAGCACCAACCGCGCGGAGACCCGACGAA..... 66996

```



```

64481 CAACGAAGCGGATCAATCCCGCGCTCCACGGTTTCCACGCACT 65432
      ::::: |||||
789 .....SerValSerAspLysAspLysArgLys 797
      ||:::|||||
65431 CCAGCAGCAACCGCTGCTGGGATCCACCAACCGCTCCAGCGGGA 65382
      |||||
798 ...LysAspLysProLysValLysSerAspLysThrLysArgLys 811
      |||||
65381 ATCCCGAAGAACCGGCATCGAACTCTGCCGATCTCGCAGGAAGCCACC 65332
      |||||
811 ..... 811
65331 CTCACACATAGAGGTCCCAAGACGATCGGATCGGATCGAACACC 65282
812 .....SerAspLysSerAlaThrAlaLysLys 820
      ||:::|||||
65281 CCGCTGATCCACGACGATCGTCCGGAATCCCGCACCGCATCCAG 65232
821 AspAsnValLeuLysProSer 827
      ||:::|||||
65231 CCCCAGGCGACCAACGCGCAAGCTCTCCGCGAGAGACACCCACCCG 65182
828 ...Lys.GlyProGlnGlnLysValAspLysArgGlnLysSerPro 842
      ||:::|||||
65181 ATACGACAGCCCATCCCGCATCACACGGATCTCAAGATCACCG 65132
843 ArgSerGlnProLeuLys 849
      |||||
65131 CCGCCACAGCGCGGTGGGAGTCCGAGCTCCACGCCACCAATCAGT 65082
850 ...LysAlaLysGlnGlnAlaThrLysIleAspSerValLysProSer 865
      ||:::|||||
65081 TTGTGAGCAGAACCGAACCAAG....CCGCGGTGTCCGAAATGTC 65038
865 SerSerGlnLysAspGlnLysValThrGlnThrProAlaLysAlaIle 881
      ||:::|||||
65037 GAACACCGCGAGTGGCGGCAACGAGCCGACACCCGCGCAAGCCGT 64988
882 SerLysSerAlaLysAspThrArgArgGlnSer 892
      |||||
64987 TCCGAATCTCCACACGACGACGAAATCAAAACCAATCCCGGAAGCA 64938
893 .....GlnProArgThrArgArgSerLysArgThrValProLys 905
      |||||
64937 CGATCGGCCCGACCTCGACCGACCCGAATGCCCGACCGCGGAAC 64888
906 .....Thr 906
64887 ATGCTCGCGCACCAACCCAAAGCTGGTTCTCTGATCCAGGTCGCGCA 64838
907 SerSerGlnLysSerGlnProValArg 919
      |||||
64837 GCCTCCGACCGACCAACGAGACCGTTCGCGCGGCTCGCTGATCC 64788
919 ArgSerLeuArgLysIleAsn.TyrLeuIleAlaArgGlnLysAsnGln 935
      |||||
64787 CGCTCCGCGCGACGGCGCTGGCACCAAGTCCGCGACACAGGAGCAC 64738
936 ArgGlnLysArgLys.....LysSerValAspLysAspPheGln 949
      ||:::|||||
64737 ACCGCGCTCGCGCATCGCGCGACGGGCTTCAGGTCCAGCGAGCG 64688
949 rSerSerMetLysIleSerLysValGlnGlnThrGlnIleValLysPro 966
      |||||
64687 GAACGAGAACGATTCGCGCGGAAACGGAACGCGATCTCAACAGCGC 64638
966 SerProLysArgLysMetGlnLysValGlnLysLeuGlnArgThrPro 982
      |||||
64637 AACCGCTCGCGCGCATCGCGGATTCGT 64606
983 GlnLysAspLysIleAlaSerSerThrThrProAlaLysLysIleLys 999
      |||||
64605 .....GCGGCATTCGCGGTACGGCCCGCGCATCGAGC 64571
999 uAsnArgLysThrGlnLysLysIleGlnLysAlaIleAsnAlaSerThr 1016
      ||:::|||||
64570 CTCGCGTCAATCCGCTGGGCTCTCCACAAAGCCCAAGCAGCAGACC 64521
1016 hTrpSerGlnProSerGlnLysLeuGlnSerThrSerSerLysIleLysGln 1032
      |||||
64520 GCAAGCAGCCC.....CGCGTCCCGGCTGCTGAAAGCCAGCGCTCCAG 64477
1033 GlnLysValLysGlnLysAla.....LysArgLysValAla 1044
      |||||
64476 GAACGCGTTCGCGCGCGGCTAGTTCCCTGACCGGACCGCCGCGCAC 64427
1044 agLysSerGlnGlnSer..SerSerThrLeuValAspTyrThrSerThr 1060
      |||||
64426 CCGAAGACGAGAGAGACGACGACGACGAGATCCAGATCAAGATC 64377
1060 rSerThrGlnLysSerProValArgLysSerGlnLysThrAspThr 1077
      |||||
64376 AACTCTGCAAGTGCACGACGCGGCGCTTCGCGCGCACACTTTCG 64327
1077 yAspThrValIleLysThrMetGlnLysThrAsnAspAsnThrAla 1093
      |||||
64326 CAGGTGCTCGGTGTC.....AGCAGCTCGAGAC 64296
1094 ProAlaGlnAspValIleIleMetIleGlnValProGlnSerLysThrAs 1110
      |||||
64295 CCGTCCGCAAGTCCCGCGCGCATCGACCACTCCGCGAGCGGAATC 64246
1110 pLysAspAspPheGlnSerGlnLysLysAsp.....ValLysThr 1124
      |||||
64245 TCCGAGATCTGCTGTCAGACGCGGCTCAACTGGCCCGATCCCGCAT 64196
1124 hGlnProIleGlnSerValGlnLysPro.....SerSerIleIle 1137
      |||||
64195 CGCAGCTCGCACCTGACCGCAGCTCCAGGCCACCACTATCCACC 64146
1138 LysAsnValThr.....ThrLysProSerAlaThrAla 1148
      |||||
64145 AGCTGCGTACCCCGGGGCGCTCCACCGCGCTCGCCCGCGAGCAC 64096
1148 aLysThrThrGlnLysGlnSerGln..... 1157
      |||||
64095 CCGCGGACACCGCGCTCCCGCACCAATCCCGGACCAACCCCGCCA 64046
1158 .....ProGlnLysLeuGlnLysLeuProLysGlnAlaSerHis 1170
      |||||
64045 GCAAGCCCGTACCGCGGAATCAACACGTCGA..... 64011
1171 GlnLeuMetGlnIleGlnLeuArgSerSerLysGlnLysAlaSerSer 1187
      |||||
64010 .....TCCGACCGGCGCATCGGATCCGCT 63985
1187 uLysGlnArgAlaLysAspArgGlnHisSerGlnLysAspAsn 1204
      |||||
63984 GTCTCCGCGCGACACATCGCGCAGGAGGAGACC...CGGACTTCAC 63938
1204 roAspLysArgLysSerGlnAlaGlnProAspLys..... 1215
      |||||
63937 CGCGGCGCAACGCGCTGCGGTCTCCGGAACGACAGCGCGCGCGAGC 63888
1216 .....GlnSerThrValAspArgLeuSerGlnGlnHisPheLys 1229
      |||||
63887 GCACGCCATATTCGGCGGTCTCCACATCGCAACACAGAAATCCGTC 63838
1229 sThrLeuSerGlnSerSerLysGln..ThrArgThrSerGlnLysHisGln 1245
      |||||
63837 CGGCTTCTCCAGTCCGCGCAGACGACAAACCCAGAGCAGACATCA 63788
1246 SerValArgGlnSerSerAsnLysAspPheThr..... 1256
      |||||
63787 CAAGTCCGAACACCTCGCGGAGCTGTGACGCGCGACACACGCTC 63738

```

```

1257 .....ProGlyArgAspLysLysValAspTyrAspSerArgAspTyrS 1271
      |||||::: :::::
63737 ACAACCAACGAGGAGATCCGGGAATCGCGCCGCCCAACCAATCCTG 63688
1271 erSerSer.....LysArgArg 1276
      |||||:::
63687 GATCAGGTCAAGCATGGTTGTGTGCGGCGACGTGCTCGGAACACGCT 63638
1277 AspGluArgGlyGluLeuAlaArgArgLysAspSerProArg..... 1291
      ::: |||||:::
63637 CCAAGTCTCTGGCGAGCGACCGCATG.....GCACACACAGCAGCTCC 63594
1292 .....GlyLysGluSerLeuSerGly..... 1298
      |||||:::
63593 GGGGCGGCGTCAAGCAACCGGTCAGTCTCGGGTGAGATTCCGTGCCAAG 63544
1299 .....GlnLysSerLysLeuArgGluGluArgAspLeuProLys 1311
      ::: |||||:::
63543 TCCGAGCCCGAGGCTCGGGGAGTGTCTATGTCGATCCTCCATCCAGG 63494
1312 LysGlyAlaGluSerLysLys.....SerAsnSerSerProArg 1325
      ::: |||||:::
63493 AGACCGGCTTGGCAGCCCGCGGCGAGATGCTTCCAGTCGACCCGGAC 63444
1325 GAspLysLysProHisAspHisLysAlaProTyrGluThrLysArgProC 1342
      ::: |||||:::
63443 AGGAGCTGCTTCATCGCAC...GGGCGCTCCCAACTGCTGTCTCGA 63397
1342 ysgGluGluThrLysProValAspLys.....AsnSer 1352
      ::: |||||:::
63396 AATGTGGCGCGCACCGCATGATGAGAGACAGCGCTTGGCCAGTCG 63347
1353 GlyLys GlnArgGluLysHisAlaAlaGluAlaArgAsnGlyLysGluS 1369
      ::: |||||:::
63346 TGTGCGTCTTAACACCGACACCGCATTCGGCGCGCGGAGAACGATCG 63297
1369 erSerGlyAlaAsnGlyHisValTyrLeuThrArgGlnThrLeuProTyr 1385
      ::: |||||:::
63296 ACCGAAGGAAGATGCGCGCTTGTGAGGACACACCTTCGCACGC 63247
1386 ArgArg 1387
      ::: |||||
63246 GAACGG 63241
seq_name: /cogn2_6/plodata/1/lna/68_COMB.seq:us-09-428-517-1
seq_documentation_block:
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428.517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

```

```

alignment_scores:
  Quality: 357.50      Length: 1622
  Ratio: 0.520        Gaps: 77
  Percent Similarity: 42.355  Percent Identity: 19.975
alignment_block:
US-09-811-045a-1 x US-09-428-517-1/rev ..
Align seg 1/1 to reverse of: US-09-428-517-1 from: 1 to: 50937
96 HistHrcysProThrcysHisGlnAsnAsp..ValSerProAspAlaLeu 111
   |||||::: |||||:::
32025 CACAACCGGCCAACATCTCATCGGCTGAGCTCAGCAGACACG 31976
112 IleAlaAsnLysPheLeuArgGlnAlaValAsnAspPheLysAsn.... 126
   ::: |||||:::
31975 GTGAGAGAGCGGCCCATYACCGACCGACCGACACCGCGCTCGAATC 31926
127 .....GluThrGlyThrLysArgLeuArgLysGlnLeuP 139
   |||||::: |||||:::
31925 CTCACGATGTGTGCGACACGCGCGCGAGCTGACGAGACATGC 31876
139 roProPheLeuPheLeuValProProProArgProLeuSerGlnArgAsn 155
   |||||::: |||||:::
31875 CGCCC.....CCACACGAGATCCGCGGATCCGCGGCTC 31841
156 LeuGlnProArgSerArgSerProIleLeuArgGlnGlnAspProValVa 172
   ::: |||||:::
31840 CGCAACACACACACCGCCGCGCATGCGCGCATCAACCCCGACGCAC 31791
172 lPheArgTyrThrValSerProThrcysSerAspThrLysThrAlaLys 189
   ::: |||||:::
31790 ACAAGCGGACGATCTCACCCTCGAATGCCCAACACGACACAGCT 31741
189 er..... 189
   |||||:::
31740 CCACACCGTAAATACCGCCAGGTCGAGCGCATGACACCATCAGCCGCAC 31691
190 .....CysSerAspSer.....GlyThrLeuSerArgLeuP 200
   |||||::: |||||:::
31690 AACGCGGCTGACCATCAACCCGCCCAAGAACAGTGGCTCCCTCC 31641
200 Ala.....ProSerIleSerSerLeuThrSera 210
   |||||::: |||||:::
31640 CTGCACACCTCGACACGACCAACCGCTCAGCGGATTCAGACCGCAG 31591
210 snGlnSerSerLeuAlaProProValSerGlyAsnProSerSerAlaPro 226
   ::: |||||:::
31590 CGCACTCGGCAACCGCTCCGCGAAGCGGACGCAATCCAGCAACCCA 31541
227 AlaProValProAspIleThrAlaThrValSerIleSerValHisSerG1 243
   |||||::: |||||:::
31540 GCCGCCATCC.....ACCCACTGCGGA 31518
243 LysSerAspGlyProPheArgAspSerAspAsnLysLeuLeuProAlaA 260
   |||||::: |||||:::
31517 ACCCTGACACAGA.....AAGACGACACGACGAGACGACCTCAG 31477
260 lAlaLeuThrSerGlnHisSerLysGlyAlaSer..SerIleAlaIleTh 276
   |||||::: |||||:::
31476 CCGGAGCCACACCGACACCGCCGCGACACCGCCAC..... 31438
276 rAlaLeuMetGluLysGlyValProGlyThrSerProThrPasnSerI 293
   ::: |||||:::
31437 .....CCGACGCAAGCGGCGCAACGCAACCGCGCATGAGTCCA. 31399
293 lPheValGlyLysLeuLeuHisGlyGlnLeuIleProThrThrGly 309
31399 ..... 31399
310 ProValArgIleAsnAlaAlaArgProGlyGlyArgPro..GlyTyr 325
   |||||::: |||||:::

```

```

31398 ... GTTCACTGCCGATCGCAACAGCCCGGTGCTCGAAGACCGACCTCTGTC 31352
325 pgiuHisSerAsnLysLeuGlyTyrLeuValSerProProGlnGlnIleA 342
31351 G.....CCACCAATG..... 31342
342 rgarGlyGluArgSerCysTyrArgSerIleAsnArgGlyArgHisHis 338
31341 .....ACCAGCCCATCATCGACCGCATTAAC 31315
359 SerGluArgSerGlnArgThrGlnSerProSerLeuPro.....Al 372
31314 CCGACACGTCGCGAG.....CCAAAGCTGCGCCTTGACAGC 31277
372 atPrProCysPheValProValProProProLeuTyrProPro.... 387
31276 AAGCCCTCGCAT...CCGCGCGAGACACCCACGCGACACAGATC 31230
388 .....ProProHisThrLeuProLeuPro.....ProGlyVal 398
31229 ACCGGACGACACACAGAGTTCGCGCTCGGCTCGGTTCCGAGGCG 31180
399 ProProProGln..... 402
31179 CTTCTCTCAAGATCACATGCGCGTCTGTCGACTGACACCGAAGCGGAG 31130
403 .....PheSerProGlnPhePro..... 408
31129 ATGCGTGGCCGAGAGAGAGCGCTGAGCTCGGCTCCAGCGCGCTCTTC 31080
409 ..SerSerGlnProPro..... 413
31079 AGTGAAGAACTCCACCGCAGCCGCGCATCGCTCCAGCGCGCTCTTC 31030
414 .....ThrAlaGlyTyrSerValProProProGly.. 423
31029 CATCCACATGCAAGCTTCTGGGACACACCGCCCGACGCGCAGAGACC 30980
424 .....PheProProAlaProAlaAsnIle 432
30979 ATCTTCATCACACGACACACCGCGCGCTGATGATGATGATGATGATGAT 30930
432 ePrHisLysPheSerProGlyValProThrAlaHis..... 444
30929 CGAAGTTTACCG...ACCCAGACCAACCGGAGCGCCACCGCGACGCCCT 30883
445 SerAsnThrMetProThrThrGlnAlaProLeuLeuSerArgGluGluPhe 461
30882 GCCCATACGTGGCAACACGCGCTGAGCCT..... 30853
461 eTyrArgGluGlnAsnAspLysGlyArgLysLysPheProTyrSerG 478
30852 .....CGATCGATCACCCCAACCGCGCTGATGATGATGATGATGATGAT 30816
478 LysSerSerTyrSerArgSerSerTyrThrAspSerSerGlnGlyLeuAla 494
30815 CTCACACCAATCCACATCAGCAGAGGACGACCGCGCTGACGCAACCGCG 30766
495 GlnHisIleHisAlaLeuThrLeuSerProSerAlaAlaHisThrLeuAs 511
30765 CAGGAGTCAACCGCTGCTGGAGGAGGACATTCGCGCGCGTCAACCCATTC 30716
511 pLeuLeuHis...AspHisProHisProProGluGluAlaGluAlaArgS 527
30715 GAGCGACCATCTGATGACCGCATCCCGGACCA..... 30679
527 ePrAlaMetIleValHisMetProAspLeuMetAspIleAlaHisAla.Ar 543
30678 .....CCGCGCAACACCGCATGACCCCAACCGCGCGC 30649
543 gSerArgSerProProTyrArgArgTyr..... 552
30648 CATCCGCAACCGCTGCAACCAACACACACACCGCTGCGGACGACCA 30599

```

```

553 .....ArgSerArgSerArgSerProProGlnPheArgGlyGln 565
30598 AAMCATCCGCGGAGCGGGAACGCTTGACACCGCCATCCACGCAAA 30549
566 SerProThrLysArgAsnValProArgGluGluGlyGluGlu..... 580
30548 ACCCG.....CTGCCGACTGAACGTCTCAACGTCAACGCGCGC 30511
581 .....TyrPheAsnArgTyrArgGluValProPro 591
30510 ACGACATCACCGCTCACCCACCCACCAACGCGGACGCAACCTACCCCGC 30461
591 rOTyrAspIleLysAlaTyrTyrGlyArgSerValAspPheArgAspPro 607
30460 CG.....CAACCC 30453
608 PheGluLysGluArgTyrArgGluTyrPGLuArgLysTyrArgGluTyrP 624
30452 CTGCACCGCCGATGCAACGCGCAACAGCAGCAGCAACGCGCGATCCA 30403
624 r..... 624
30402 CCGTCAACCGCAGGCGCCCTCAAGACCGATGAGCAACACGCGCGAA 30353
625 .....GluLysTyrTyrLysGlyTyr... 631
30352 GCGCGCCCTGATGTGTGCGCATGAGCGAGCTGCGCTTCCAGCGCTTCGAC 30303
632 ..AlaValGlyAlaGlnProArgProSerAlaAsnArgGluAsp.PheS 647
30302 CAGCGGTAGGGCTGCGCGCATAGCCGCTGCTCCACGCGCCATGTACA 30253
647 ePrProGluArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArg 663
30252 CGCGCGGTGCTGCTGCTGCTGAGAGCGCGCGCTTACACCGGCGCTCTCG 30203
664 GlyArgArgGluAspTyrAlaAlaGlyGlnSerHisArgAsnArgAsnLe 680
30202 AATGCTCTCCAGAGACGTC.....TCCAGCAGCAACCGCGCTCTG 30165
680 uGly.....GlyAsnTyrProGluLysLeuSer 690
30164 CCGGTCCATGCCACAGCCCTCACCGCGGCGAGATCCGAAAGACGCGGCT 30115
690 hArgAspSerHisAsnAlaLysAsnProLysSerLysGluLysGlu 706
30114 CG.....AACTCCCGGAGTCTACAGAAACCGCGCTCACGC 30077
707 SerGluAsnValProGly...AspGlyLysGlyAsnLysHisLysLysHis 722
30076 ACPATACAGCTACCGCGCGCTCGGATCGGATCG.....TA 30039
722 sArgLysArgArgAsnGluGluLysGlyGluGluSerGluSerPheLeuA 739
30038 CAGGCGCCCAAGTCCACACCGCGGCTCTGCGGTACACCGCGATGCCCT 29989
739 snProGluLeuLeuGluThrSerArgLysCysArgGlySerSerGlyIle 755
29988 CACGCGCGCTGTCAGCAGAGTCCACAGAACTGTGGCGGTGAGCGGTG 29939
756 AspGluThrLysThrAspThrLeuPheValLeuProSerArgAspAspAl 772
29938 GCGTGGCGCGCGCGGAGGACCGGACGCTCATTCACAGTGGCGATCGG 29889
772 atPrProValArgAspGluProMetAspAlaGluSerIleThrPheLys 789
29888 CTCTCGGCGCGCTTCTGCTGCGGAGTCAATCGATCGTCACTCTCTGA 29839
789 ePrValSerAspLysAspArgGluLysAspLysProLysValLysSer 805
29838 GTTCGCTGCTGGCGCGCAACAGATATTCGCGCACTTCTCGCGCTACGC 29789

```

[illegible]

| | | |
|-------|---|-------|
| 28930 | TCGGTGGCATGGGCGCATGCCCGCGGCTCAATCCGGTACGGTGCAT | 28881 |
| 1063 | | 1063 |
| 28880 | GTCCGCGACAGCCCGCGTGAGACCCGTGGCCCTCGTCCACACCCCGCAGC | 28831 |
| 1063 | | 1063 |
| 28830 | CGAGTGCACAGGAGGTGACCCCGCGCGCCGGGTGTCCGCGAGGGTG | 28781 |
| 1064 |GlySerProVa 1067 | |
| 28780 | TCCAGACTGATTGGCCGCGCGTAACCTGCCCCGGCGCGCCCGCCG | 28731 |
| 1067 | lArgLysSerGIuGIuLysThrAspThrLysArgThrValIleLysThrm | 1084 |
| 28730 | CATGCTCGCTGCCCGATGAGAAATCAGACAGGCGCGGGTGCAGAGTCCA | 28681 |
| 1084 | eGIuGIuLysAsnAsnAspAsnThrAlaProAlaGIuAspValIleIle | 1100 |
| 28680 | GTTCGGCGCATCACTGACCTCAGAGGTACAGGCGCGCGTCCACCTGGGTTTG | 28631 |
| 1101 | MetIleGIuValProGIuSerLysTrpAspLysAspAspPheGIuSerGI | 1117 |
| 28630 | AGACACGTGTCCACCCGTTGACAGGGGTGAGACCCGCGACACACCTCGTC | 28581 |
| 1117 | uGIuGIuAspValLysThrThGIProIle..... 1127 | |
| 28580 | GAGCACGCCGCGCGTGTACAGCATGCGCTGTGACCGAGTCTCGCGCGTA | 28531 |
| 1128 |GlnSerValGIuLysProSerSerIleIle 1137 | |
| 28530 | TGCTCTCCAGAGAGGCGCTTGAGCGCGCTTGCGGTGGCGACGTCCAGAGCC | 28481 |
| 1138 | LysAsnValThrThrLysProSerAlaThrAlaLysTYrThrGIuLysGI | 1154 |
| 28480 | GCCACCGTACCTCGCGCGCCGCGCCACCGACCGAGTTGG...GCCGCAAGTTC | 28434 |
| 1154 | uSerGIuGIuProGIuLysIleu..GIuLysIleuProLysGIuAlaSerH | 1170 |
| 28433 | GGGGGCGACGCGCGTGGCGGTGTACAGACGACGCGCGCGAGCTC.... | 28389 |
| 1170 | ScIuIuMetGIuGIuLysIleuArgSerSerLysGIuSerAlaSerSerg | 1187 |
| 28388 | ...GTGCTCGGTGACAGAGTGGCGGGGACCAAGGGGTCGCCAGCTGCC | 28343 |
| 1187 | IuLysGIu.....ArgAlaLysAspArg 1194 | |
| 28342 | GCCCCGTGTATCAGCACGGTTCCTCGGCGTTCGACGCGCGCGGACGGT | 28293 |
| 1195 | GIu.....HisSerLysSerGIuLysAspAsnProLysLysAr | 1207 |
| 28292 | GAGGAGACACTTGGCCACATGACGTGCTCGGCTGTGACGACCTCAATGG..CC | 28244 |
| 1207 | GIuSerGIuAlaGIuProAspLysGIuSerThrValaLysArgLysSerg | 1224 |
| 28243 | TGCTGGGCGCGCGGACCGGCGGACCGCGGCAAGCGCG.....GCMA | 28203 |
| 1224 | IuGIuGIuHisPheLysThrLysSergIu...SerSerLysGIuThrArg | 1239 |
| 28202 | CCGCAAGGACACCGGCGTGCAGAGCGCCATGATCTTCGTGACACATCTGCC | 28153 |
| 1240 | ThrSerGIuLysHisGIuSerAlaIArgLysSerSerAsnLysAspPheTh | 1256 |
| 28152 | CGATCGCTCGGCGACCGCGCTCGGCGAGGTGCGAGCGACGTACCGGATG | 28103 |
| 1256 | rProGIu.....ArgAspLysLysValaLysPyrLysAspSerArg | 1269 |
| 28102 | TCCGGGACGCCGTCCGCCCGGACGACGCCGAGTGGGATACGATTCGGT | 28053 |
| 1269 | sPyrLysSerSerLysArg...ArgAspGIuArgGIuGIuLysIleuAlaArg | 1284 |


```

28052 CTTGCCATCTGCAGAACCGCCGCTGCGCAGAGACGCGCGAG 28003
1285 ArgLysAspSerProAlaGlyLysGluSer..... 1295
28002 CGTCGACGAGATGACCCGAGAGGCGATTCAGAGACATGATCCGCGC
1296 .....LeuSerGlyGlnLysSerLysLeuArgGlu.....G 1306
27952 CCACACATGCTCCGCGGAGACCGCTGCTGCACTGGTCTGTCGACGA
1306 LuArgAspLeuProLysLysGlyAlaGluSerLysSerAsnSerSer 1322
27902 GCGGATGTCCTGCTGTCGAGAGCCGTCGCGCGAGACATCCACTTGC
1323 Pro.....ProArgAs 1326
27852 CCTTGTCTGCGGTCGCGACACCTCGGCATCCAGTCCGCTGCCAATGC
1326 PLYSLYSProHISAspHISLys.....AlaProTyrGluThrL 1339
27802 ACGGCGGCCATGCGCGACACCGCCGACGCGAGTGCACACGACGACTC
1339 ysArg..... 1340
27752 GCCGCGCGCGAGGCTGTCAGTCTGCAGGCGGTAGTCGCGGTACGA
1341 ..ProCysGluGluThrLysProValAspLysAsnSerLysGluAr 1356
27702 ACGCCACCGCGACCGACGCGCGCT.....CGCGCAG 27671
1356 gGluYsHISAlaAlaGluAlaArgAsnGlyLysGluSerSerGlyAla 1373
27670 GACCGACGACGCGGATCGCTGACGA.....GCATCCGGTGGTGGT
1373 sn 1373
27629 CA 27628
seq_name: /cgn2.6/ptodata/1/lna/5A.COMB.seq:US-08-320-559-1
seq_documentation_block:
: Sequence 1, Application US/08320559
: Patent No. 563135
: GENERAL INFORMATION:
: APPLICANT: Croce, Carlo
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
: TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
: TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
: TITLE OF INVENTION: All-1 Region
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 563135rlis
: STREET: One Liberty Place - 46th Floor
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/320,559
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/062,443
: FILING DATE: 14 MAY 1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/971,094
: FILING DATE: 30-OCT-92

```

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/888,830
: FILING DATE: 27-MAY-92
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/805,093
: FILING DATE: 11-DEC-91
: ATTORNEY/AGENT INFORMATION:
: NAME: Deluca, Mark
: REGISTRATION NUMBER: 33,229
: REFERENCE/DOCKET NUMBER: TUD-0855
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1425
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: ANTI-SENSE: NO
: US-08-320-559-1

```

```

alignment_scores:
Quality: 355.00 Length: 1489
Ratio: 0.590 Gaps: 64
Percent Similarity: 40.430 Percent Identity: 19.946

```

alignment_block:

US-09-811-045A-1 x US-08-320-559-1 ..

Align seg 1/1 to: US-08-320-559-1 from: 1 to: 14255

```

90 LeuLeuGluSerAspLysHISThrCysProThrCysHISGlnAsnAspVa 106
892 ATTATTCCTCTTCAAAAAGACA..... 915
106 LserProAspAlaLeuLeuAlaAsnLysPheLeuArgGlnAlaValAsn 123
916 .....GATGCAACCATTCGTAAACATCTTACACAGGACA..... 951
123 snPheLysAsnGluThrGlyTyrThrLysArgLeuArgLysGlnLeuPro 139
952 .....AAAAGGGGCTCAAAAAGAAATGGAAGAAAGACGCT 990
140 ProPhe.....LeuPh 143
991 CAGCTGACGAGGAGAAAGTGAAGACACAGTCAAAAATATTCGACAGTT 1040
143 eLeuValProProAlaProAlaProLeuSerGlnArgAsnLeuGlnProArgS 160
1041 CATCATCCTGTCGTGTCAGTCTATCTCTCGCGGATCATTAAGACCCCTC 1090
160 eArgSerProLeuArgGlnGlnAspProVal..... 172
1091 GCGCGTTTATAGAGATGAGATATATGACCCCTCAATTAATGCCCCGA 1140
173 .....PheArgTyrThrValSerProThrCysSerAs 183
1141 TTAGAGTCTACACCGANTAGATGATGAGT...GCCCGTCGTGTGATC 1187
183 pThrLysThrAlaGlySerCysSerAsp.....SerLysT 195
1188 TTCTGAAAATAATGACGTCACCTCTCGACCTCCTCAATATGCTTCA 1237
195 hLeuSerArgLeuProAlaProSerLieserSerLeuThrSerAsnGln 211
1238 ACTCCTCTCGATCTAGTAGCCCGCAGTGTGATATCTTCAACAGACTCTCAG 1287
212 SerSer.....LeuAlaProProValSerGlyAsnProse 223
1288 GCTTCTGAGGAGATTCAGTACTTCTGAGAGAGCGGCGCATACCCCTGA 1337

```

```

223 rSerAlaProAlaProValProAspIleThrAlaThrValSerIleSerV 240
    : ||| ||| ||| |||
1338 AGTTTCATCT...CCACTGCC...ATTCCAGTCCC 1369
240 aHisSerGluLysSerAspGlyProPheArgAspSerAspAsnLysLeu 256
    : ||| ||| ||| |||
1370 CAGAAATGAGATTAATGATAGGAAAGCATATCA..... 1410
257 LeuProAlaIleAlaLeuThrSerGluHisSerLysGlyAlaSerSerI 273
    : ||| ||| ||| |||
1411 .....GTGTGGAGAGAAAGTTTGGATCTAAGC.. 1440
273 eaIleIleThrAlaLeuMetGluLysGlyValProGlyThrSerProT 290
1440 ..... 1440
290 rPasnSerIlePheValGlyInSerLeuHisGlyGlnLeuIlePro 306
1440 ..... 1440
307 ThrThrGlyProValArgIleAsnAlaAlaArgProGlyGlyArgP 323
1440 ..... 1440
323 oGlyTrpGluHisSerAsnLysLeuGlyTyrLeuValSerProProGln 340
    : ||| ||| ||| |||
1441 .....ACGAAAAATTATCAACTCTACAAAGTCCCCCAG. 1476
340 InIleArgArgGlyGluArgSerCysTyrArgSerIleAsnArgGlyArg 356
1476 ..... 1476
357 HisHisSerGluArgSerGlnArgThrGlnSerProSerLeuProAlaTh 373
    : ||| ||| ||| |||
1477 .....CAGAGACCTCTCGTCT..... 1494
373 rProCysPheValProValProProProLeuTyrProProProProH 390
    : ||| ||| ||| |||
1495 .....CCACTCCACTCTGCTGACTCCACCGCCA. 1524
390 IsthIleuProLeuProProGly..... 397
1525 .....CCACTGCAGCCAGCTCCACTATCTGACCACACACCTTGG 1566
398 ..ValProPro.....ProGlnPheSerProGlnPheProSerSerG 411
    : ||| ||| ||| |||
1567 CTATGCTCCCAACAATCCCTTAGCATCACATTTTGGCTTCCAC 1616
411 nProProThrAlaGlyTyrSer.....ValProProProGlyPhe. 424
    : ||| ||| ||| |||
1617 TGCTTCCTATGCAAGGGAAGCGAAATCTATTTCGAGAACCGACATTTA 1666
425 .....ProProAlaProAlaAsnIleSerThr 433
1667 GGTGGACTTCTTTAAAGCATCTAGTCAAGCCCAACTACTTTCTTCA 1716
434 AlaCysPheSer...ProGlyVal.....ProThrAlaHisSerAsnTh 447
    : ||| ||| ||| |||
1717 GCAAGATGCAAGAGAGGTCTTATTCGCAACCAATATTGATAATTT 1766
447 rMetProThrThrGlnAlaProLeuLeuSerArgGluIlePheTyrArg 464
    : ||| ||| ||| |||
1767 CCGACCCCTCCACTACTCC..... 1788
464 LuginAsnAspLysGlyArgGluSerLysPheProTyrSerLysSer 480
    : ||| ||| ||| |||
1789 .....GAGACGCTTGGCTTGCATCTGTTTGCATCTGGTACCGCT 1833
481 TyrSerArgSerSerTyrThrAspSerSerGlnGlyLeuAlaGlnHisI 497
    : ||| ||| ||| |||
1834 GCTTAGCCCGATTTGTTTGGCCACTCATCTTGGAACAAGTTTGATAT 1883
497 eHisAlaLeuThrIleuSerProSerAlaAlaHisThrLeuAspLeuH 514

```

```

1884 GCAC.....AAAAGAGCCCT.....CTCTGCA 1906
    : ||| ||| ||| |||
514 IsAspHisProHisProProGluGluAlaGluAlaArgSerAlaMetIle 530
    : ||| ||| ||| |||
1907 GAGCTCCAAATTTTACTCCAAAGTAGGCTCATCTGAAATATTGAGTCT 1956
531 ValHisMetProAspLeuMetAspIleAlaHisAlaArgSerArgSerP 547
    : ||| ||| ||| |||
1957 GTAACTTGCCTAGTAAATTCGAACTTCTGTGAACACATCTTCTCAGAGT 2006
547 oProTyrArgArgTyrArgSerArgSerArgSerProProGluPheArg 564
    : ||| ||| ||| |||
2007 ATCCAAATGAAAAAGAAAAAGAAAGTTTATGCTTATTCGATTCGAC 2056
564 LysInSerProThrLysArgAsnValProArgGluLysGlyArgGlu 580
    : ||| ||| ||| |||
2057 CAAATCTCTCT...CACTCCATGAGACCAAGATGGAAGCTT 2100
581 TyrPheAsnArgTyrArgGluValProProProTyrArgIleLysAlaTy 597
    : ||| ||| ||| |||
2101 AGTAGTCTGAGCTTCACCTCACCCGCCCGTCTGCTCTCTCTC 2150
597 rTyrGlyArgSerValAsp.....P 604
2151 GTTAGCATTTCTGTAGTCTCTTGCCTAGTGCCTTAACCCACTT 2200
604 hArgAspProPheGluLysGluArgTyrArgLutProGluArgLysTyr 620
    : ||| ||| ||| |||
2201 TTACTTTCTCTCT..... 2214
621 ArgGluTrpTyrGluLysTyrTyrLysGlyTyrAlaValGlyAlaGlnP 637
    : ||| ||| ||| |||
2215 .....CAFTCCGACTCAGTCTGGGAATCTGCAGAAAAATCA 2255
637 oArgPro.....SerAlaAsnArgGluAspPhe.....S 647
    : ||| ||| ||| |||
2256 GAGACCAAGAGACAGACTACTGCTCGGCGAGACCAATTTCATCAAGTA 2305
647 eProGluArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArg 663
    : ||| ||| ||| |||
2306 GTCTACTCTCTCTCTCTGTTTACCCAGGCTCTCAGACTGAAAGA 2355
664 GlyArgArgGluAspTyrAlaAlaGlyGlnSerHisArgAsnArgAsnLe 680
    : ||| ||| ||| |||
2356 GGGAGAAATTAAGACAAAGCC..... 2376
680 uGlyLysnTyrProGluLysLeuSer...ThrArgAspSerHisAsnA 696
    : ||| ||| ||| |||
2377 .....CCGAGAGAGTGTCCAAAGATCGAGATGCTGCAAGA 2413
696 IaLysAspAsnProLysSerLysGluLysGluSerGluAsnValProGly 712
    : ||| ||| ||| |||
2414 GCGTGCAGAAAGACAAAGTAGAGAGAGACCGGAGAGAGA..... 2454
713 AspGlyLysGlyAsnLysHisLysLysHisArgLysArgAsnGluGlu 729
    : ||| ||| ||| |||
2455 ..GAAAGAGAAATAGCGGAGTCAAGAGAAAGAAAGAAAGG 2501
729 uLysGlyGluGluSerGluSerPheLeuAsnPro..... 740
    : ||| ||| ||| |||
2502 ATCAGAAATTCAGACTACTTCTGCTTGTATCTCTGTGGTAGGGTTTCCA 2551
741 .....GluLeuLeuGluThrSerArgLysCysArgGly 751
    : ||| ||| ||| |||
2552 AAGAGAGGTTGTTGTGAAGATGTTGCCACTTCACTTCTGCCAAAAA 2601
752 SerSerGlyIleAspGluThrLysThrAspThrLeuPheValLeuProSe 768
    : ||| ||| ||| |||
2602 GCACAGGCGGAGAAAGACTTCA.....TC 2627
768 rArgAspAspAlaThrProValArgAspGluProMetAspAlaGluSerI 785
    : ||| ||| ||| |||

```

```

2628 ACATGATTCTGGGACT.....GATATTACTTCTG 2656
785 LethrThelysSerValSerAspLysArgGluLysAspLysPro 801
    ::|||::
2657 TGACTCTT..... 2664
802 LysValLysSerAspLysThrLysArgLysSerAspLysSerAlaThrAl 818
    |||:::||||
2665 .....GGGATCAACAGC 2678
818 aLysLysAspAsnValLeu..LysProSerLysGlyProGlnGluLysV 834
    |||:::|||||
2679 TGTCAAAACCAAAATACTTATAAAGAGAGAGAAATCTGAAAAAA 2728
834 aLAspLysAspArgGluLysSerProArgSerGluProPoleLysLys 850
    ::|||:::|||||
2729 CCAACTTGAC.....CTCGGCCA...ACTGCCATCTCCGGAGAG 2769
851 AlAlysGlu.....GluAlaThrLysIleAspSerValLysProse 864
    |||:::|||||
2770 GAGAAAACCTTGCTTCCTTCCACTCTTCATCTAGCACTGTAAACATTG 2819
864 rSerSerSer.....GlnL 869
    |||:::|||||
2820 CACTCTCTCATAGGCTCCATGTTGGCTCAGCAGACAACCTTCCAAATGA 2869
869 yAspGluLysValThrGlyThrProArgLysAlaHisSerLysSerAla 885
    |||:::|||||
2870 CTGACACAGAGGTTGCCAGCTCTTAAAAAAGCCAAAGCTCAGCTGTGC 2919
886 LysAspThrArgArgInserGlnProArgThrArgArgSerLysArgTh 902
    |||:::|||||
2920 AAGATTGAGAGAGTAGAGCTTAAACAAACGACAG..... 2958
902 rValProLysThrSerSerGlnLysSerGlnProValArgThr.....A 917
    |||:::|||||
2959 ....CCAAAGCACAGGGTCACAGAAGTGACATCAGACAGCTGTGTGC 3004
917 rGArgProArgSerLeuArgLysIleAsnTyLLeuIleAlaArgGluLys 933
    |||:::|||||
3005 GAGGACCCCG.....ATTAAACAT...GTCTCAGAGAAGCA 3039
934 AsnGluArgGluLysArgLysLysSerValAspLysAspHecLysSer 950
    |||:::|||||
3040 GCTGTGCCCTTGCGGAAAAAGAGCTGTGTCTGTAT..... 3078
950 rSerMetLysIleSerLysValGluGlyThrGluIleValLysProSer 967
3078 ..... 3078
967 rLysArgLysMetGluGlyAspValGluLysLeuGluArgThrPro... 982
    |||:::|||||
3079 .....GACATGCCACCTCGAGTGCCTTACCATG 3108
983 ...GluLysAspLysIleAlaSerSerThrThrProAlaLysLysIle 998
    |||:::|||||
3109 GAAGAAGAGAAAAAGATTGTCTTC..... 3135
998 sLeuAsnArgGluThrGlyLysIleGlyAsnAlaGluAsnAlaSer 1015
    ::|||:::|||||
3136 .ATGGGGAATGATGACAAAGTCATCAATGCTGCTCGAAGATGCTGAAC 3184
1015 hrThrLysGluProSerGluLysLeuGluSer...ThrSerSerLysIle 1030
    |||:::|||||
3185 CTCTGTCTCCACCCATCAACCAATTAAACCTGTCACTAGAAACAGCA 3234
1031 LysGlnGluLys.....ValLysGlyLysAlaLysArgLysValAlaGlu 1045
    |||:::|||||
3235 CCCAGAGAACCTCCAGTAAGAAAGAGCGATCGAGCGGTGGGCA 3284
1045 ySerGluGly.....SerSerSerThrLeuValAspTyr 1057
    |||:::|||||
3285 GTGTCCCGGCTGCCAGGTGCTGAGGACTGTGTGTGTACTAATTCGT 3334

```

```

1057 hrSerThrSerSerThrGlyLysSerProValArgLysSerGluGluLys 1073
    ::|||:::|||||
3335 TAGATAGCCCAAGTTGTGGTGCATATTAAGAGAGCTGTCAGAG 3384
1074 ThrAspThrLysArgThrValIleLysThrMetGluGluTyrAsnAsn 1090
    |||:::|||||
3385 .....ATGAGAAAATCTCAGAAATCT 3404
1090 pAsnThrAlaProAlaGluAspValIleIleMetIleGlnValProGln 1107
    |||:::|||||
3405 ACAATGATGCTCTCCAAAGCTTACTCGCAGAGCAAGCTAAAGCTGTGA 3454
1107 eLysThrAspLysAspAspHecLysSerGluLysGluLysValLysThr 1123
    |||:::|||||
3455 AAAAGAAAGAGAAAGTCTAAGACCAAGTAAAGAAAGAC..... 3495
1124 ThrGlnProIleGlnSerValGlyLysProSerSerIleIleLysAsn 1140
    |||:::|||||
3496 .....ACCAAGAGACGAGTGTGTGAAGAACGT 3524
1140 LThrThrLysProSerAlaThrAlaLysTyrThrGluLysGluSerGlu 1157
    |||:::|||||
3525 GGTG.....GACTTATGTCAGA 3541
1157 LProGluLysLeuGlnLysLeuProLysGluAlaSerHisGluLeuMet 1173
    |||:::|||||
3542 AACCT..... 3546
1174 GlnHisGluLeuArgSerSerLysGlySerAlaSerSerGluLysGly 1190
3546 ..... 3546
1190 gAlaLysAspArgGluHisSerGlySerGluLysAspAsnProAspLys 1207
    ::|||:::|||||
3547 .....ACCCATCAGACAGAGAGATCTGCCCCAA 3577
1207 rGlySerSerGlyAlaGlnProAspLysGluSerThrValAspArgLeuSer 1223
    |||:::|||||
3578 AGAAAGCACTAGTAGCCCTCCACAGAAAGCCGTCGAGGAAAGAGT 3627
1224 GluGlnGlyHisPheLysThrLeuSerGlnSerSerLysGlnThrArgTh 1240
    |||:::|||||
3628 GAGAGAGGATGTCTCGGCCCTGGGCTGATTCCAACAGGCCACAC 3677
3678 TCCAGCTTCCAGAGATCAAGCAAGCAGCTCTCCAGCCAGCACTGGTCA 3727
1256 .....ThrProGlyArgAspLysLysValAspTyrAsp 1266
    |||:::|||||
3728 TCCCGCTCCAGCCACTACTACAGAGACCCGCAAGAAA..... 3765
1267 SerArgAspTyrSerSerSerLysArgArgAspGluArgGlyGluLeuAl 1283
    |||:::|||||
3766 .....CAAGTTC 3773
1283 aArgArgLysAspSerProProArgGlyLysGluSerLeuSerGlyGlnL 1300
    |||:::|||||
3774 CAAAACACACTCTAGTGAGCCCAAGAAAGACAGCT..... 3810
1300 ySerLysLeuArgGluGluArgAspLeuProLysLysGlyAlaGluSer 1316
    |||:::|||||
3811 .....CCAGCACCAAAATCAGGTCACAGACAG 3837
1317 LysLysSerAsnSerSerProProArgAspLysLysProHisAspHisLys 1333
    |||:::|||||
3838 AGCAAAACAGAAAAAGTGGCTCCCGCCCAAGTATCCCTTAACAAAA 3887
1333 sAlaProTyrGluThrLysArgProCysGluGluThrLysProValAspL 1350
    |||:::|||||
3888 A...CCAAAGAAAGGAAAGAACCA.....CTCCGCTGATA 3922

```



```

1440 ..... 1440
290 rpsasrIlePheValGlyInSerLeuHisGlyInLeuIlePro 306
1440 ..... 1440
307 ThrThrGlyProValArgIleAsnAlaIleArgProGlyGlyArgPr 323
1440 ..... 1440
323 oGIYrPGlnHisSerAsnLysLeuGlyTyrlLeuValSerProGlnG 340
1441 ..... ACGAAAAATTATCACTGTACAAAGTGGCCCCAG. 1476
340 InIleArgArgGlyArgSerCysTyrlArgSerIleAsnArgGlyArg 356
1476 ..... 1476
357 HisHisSerGluArgSerGlnArgThrGlnSerProSerLeuProAlaTh 373
1477 ..... CAGAGACCTCTCTGCT. .... 1494
373 rProCysPheValProValProProProLeuTyrlProProProH 390
1495 ..... CCACCTCCACCTGTGTGATCCACCGCCA. 1524
390 IsthIleuProLeuProProGly..... 397
1525 ..... CCACGTGACCGACCTCCAGTATCTGTGACACACACCTGG 1566
398 ... ValProPro..... ProGlnPheSerProGlnPheProSerSerG 411
1567 CTATAGCTCCAAACAATCCCTTATGACATCACTATTTGGCTGCTTCCAC 1616
411 nProProThAlaGlyTyrlSer..... ValProProProGlyPhe. 424
1617 TGCTCTATGCAAGGAAAGAAATCTATTGTGCGAAGCCGACATTTA 1666
425 ..... ProProAlaProAlaAsnIleSerThr 433
1667 GGTGACTTCTTAAAGCATTTAGTGTACAGACCAACAATCTTTCTCTCA 1716
434 AlaCysPheSer... ProGlyVal..... ProThrAlaHisSerAsnTh 447
1717 GCAAGATATGCCAAAGAGAGTCTTATTCGCAACCAATATTGTGTAATTT 1766
447 rMetProThThrGlnAlaProLeuLeuSerArgGlyGluPheTyrlArg 464
1767 CCGAGCCCTCCACTACTCC..... 1788
464 LuginAsnAspLysGlyArgGluSerLysPheProTyrlSerGlySerSer 480
1789 ..... GAGGACGTGGCTTGTGCATCTGTTTCTGTGATCTGTGACCGCT 1833
481 TyrSerArgSerSerTyrlThrAspSerSerGlnGlyLeuAlaGlnHisI 497
1834 GCTTCACCCGATGTTTTCGCCACCTCATCTTGCAGAACAGTTGTGATAT 1883
497 eHisAlaLeuThrLeuSerProSerAlaAlaHisIsthLeuAspLeuH 514
1884 GCAC..... AAAAGAGCCCT..... CTCTGA 1906
514 IAspHisProHisProGlnGluAlaGluAlaArgSerAlaMetIle 530
1907 GAGCTCCAAAGATTATCTCAAGTGAGGCTCACTGTAGAATATTGAGCT 1956
531 ValHisMetProAspLeuMetAspIleAlaHisAlaArgSerArgSerPr 547
1957 GTAACCTTGCTAGTATGCAACTTCTGCGAACAATCTTCTTCAGAGT 2006
547 oProTyrlArgArgTyrlArgSerArgArgSerProProGluPheArg 564
2007 ATCCAAATAGAAAAAGAAAAAGTGTGTAGCTTATTCGATCTGAAAC 2056

```

```

564 LysInSerProThrLysArgAsnValProArgGluGlyLysGluArgGlu 580
2057 CAAGATCTCTCT..... CACTCCATGAGACCAAGAAAGTGAAGGCTT 2100
581 TyrPheAsnArgTyrlArgGluValProProProTyrlAspIleLysAlaTy 597
2101 AGTAGTTCAGCTCTCACCTCTCACCCCCCTCTCTCTCTCTCTCTCT 2150
597 rTyrlGlyArgSerValAsp..... P 604
2151 GTTAAGCATTTCTGTATGCTCTCTGCGACATAGCTTAACCAACTT 2200
604 heArgAspProPheGlnLysGluArgTyrlArgGluTrpGluArgLysTyrl 620
2201 TTACTTTCTCTCT..... 2214
621 ArgGluTrpTyrlGluLysTyrlTyrlGlyTyrlAlaValGlyAlaGlnPr 637
2215 ..... CATTCCTGACTCAGCTCTGGGGAATCTGCAGAGAAATCA 2255
637 oArgPro..... SerAlaAsnArgGluAspPhe..... S 647
2256 GAGACCAAGAGACAGACTAGTGTCTCCGACAGACCATTTTCATCAAGTA 2305
647 eTrpGluArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArg 663
2306 GTCTCTCTCTCTCTCTCTCTGTGTTACCCAGGCTCTCAGACTGAAGA 2355
664 GlYArgArgGluAspTyrlAlaAlaGlyInSerHisArgAsnArgAsnLe 680
2356 GGGAGAAATTAAGCAAGGCC..... 2376
680 uGlyGlyAsnTyrlProGluLysLeuSer... ThrArgAspSerHisAsnA 696
2377 ..... CCCGAGAGCTGTCCAAAGATCGAGATGTGACAAAGA 2413
696 lAlysAspAsnProLysSerLysGluLysGluSerGluAsnValProGly 712
2414 GCGTGAGAGAGACACAGATGAGAGAGAGACCGGAGAGAA..... 2454
713 AspGlyLysGlyAsnLysHisLysLysHisArgLysArgArgAsnGlu 729
2455 ... GAAGAAGAGATATAGCGGAGTCAAGGAAAGAGAAAGAAAAAGG 2501
729 uLysGlyGluLysSerGluSerPheLeuAsnPro..... 740
2502 ATCAGAAATTCAGAGTAGTCTGCTTGTATCCTGTGGTAGGGTTTCCA 2551
741 ..... GluLeuLeuGluThrSerArgLysCysArgGly 751
2552 AAGAGAAAGTGTGTTGTAAGATGTGCCACTTCATCTTCTGCCAAAAA 2601
752 SerSerGlyIleAspGluThrLysThrAspThrLeuPheValLeuProSe 768
2602 GCAACAGGGGGGAAGAGTCTCA..... TC 2627
768 rArgAspAspAlaThrProValArgAspGluPrometAspAlaGluSerI 785
2628 ACATGATTCGTGGACT..... GATATTACTTCTG 2656
785 lerThrPheLysSerValSerAspLysAspLysArgGluLysAspLysPro 801
2657 TGACTCTT..... 2664
802 LysValLysSerAspLysThrLysArgLysSerAspLysSerAlaThrAl 818
2665 ..... GGGATCAACAGC 2678
818 aLysLysAspAsnValLeu... LysProSerLysGlyProGlnGluLysV 834
2679 TGTCAAAACCAAAATACTTATTAAGAAAGGAGAGAGAAATCTGTGAAAAA 2728

```

```

834 aAlaSpGlyAspArgGluLysSerProArgSerGluProProLeuLysLys 850
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2729 CCAACTYGGAC.....CTCGGCCCA...ACTGCCCATCCCTGGAGAAAG 2769
851 AlaLysGlu.....GluAlaThrLysIleAspSerValLysProSe 864
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2770 GAGAAAAACCTCTGCTTCCTTCACCTCTTCATCTAGCATGTTAAACATTC 2819
864 rSerSerSer.....GlnL 869
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2820 CACTTCCTCCATAGGCTCCATGTTGGCTCAGGACAGACAGCTTCCAAATGA 2869
869 ysAspGluLysValThrGlyThrProArgLysAlaHisSerLysSerAla 885
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2870 CTGACAGAGAGGTGGCCAGCTCTTAATAAGGCGCAAGCTCAAGCTCGC 2919
886 LysAspThrArgArgGlnSerGlnProArgThrArgArgSerLysArgTh 902
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2920 AAGATTGAGAGAGTAGAGTCTTAAACAAACGACAG..... 2958
902 rValProLysThrSerSerGlnLysSerGlnProValArgThr.....A 917
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2959 ....CCCAAGACAGAGGTCAAGAAAGTCACTCAGAGACCTCTGTC 3004
917 rGArgProArgSerLeuArgLysIleAsnTrpLeuIleAlaArgGluLys 933
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3005 GAGGACCCCG.....ATTAAACAT...GTCTGCAGAGAGCA 3039
934 AsnGluArgGluLysArgLysLysSerValAspLysAspPheGluSerSe 950
3040 GCTGTGCGCTTGGCGCGAAAGAGCTGTGTTCTGT..... 3078
950 rSerMetLysIleSerLysValGluGlyThrGluIleValLysProSerP 967
3078 ..... 3078
967 rOLysArgLysMetGluLysPValGluLysLeuGluArgThrPro... 982
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3079 .....GACATGCCACCTCGAGCTGCTTACATAGG 3108
983 ..GluLysAspLysIleAlaSerSerThrThrProAlaLysLysIleLy 998
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3109 GAAGAACGAGAAAGATTGTTCTTCC..... 3135
998 sLeuAsnArgGluThrGlyLysLysIleGlyAsnAlaGluAsnAlaSerT 1015
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3136 .ATGGGGAATGATGACAGTCAATGCTGCGTCAGAGATGCTGTAAC 3184
1015 hrThrLysGluProSerGluLysLeuGluSer..ThrSerLysIle 1030
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3185 CTCTTGTCTCCACCATCAACCAATTAAACCTGTCATCAAGAAACAGCA 3234
1031 LysGlnGluLys.....ValLysGlyLysAlaLysArgLysValAlaGl 1045
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3235 CCCCAGAACCTCCAGTAAAGAGAGCTGCATGAGGCGGTGGGCA 3284
1045 ySerGluGly.....SerSerSerThrLeuValAspTyrT 1057
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3285 GTGTCCCGCTCCAGGCTCGAGAGCTGTGTTGTACTAATTCGT 3334
1057 hrSerThrSerSerThrGlyGlySerProValArgLysSerGluLys 1073
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3335 TAGATTAAGCCCAAGTTGTGTGTCCAATATATTAAGAGAGTGTCAAG 3384
1074 ThrAspThrLysArgThrValIleLysThrMetGluGluTyrAsnAspAs 1090
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3385 .....ATGAGAAAAATGTCAGAAATC 3404
1090 pAsnThrAlaProAlaGluAspValIleIleMetIleGlnValProGln 1107
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3405 ACAATGAGTCTCTCCAAAGCTTACCTGCAGAGAGCAAGCTAAAGCTGTA 3454
1107 erLysTrpAspLysAspAspPheGluSerGluGluLysValLysThr 1123

```

```

3455 AANAAGAGAGAAAAAGTCTAAGACCAAGTGAAGAAAGAC..... 3495
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1124 ThrGlnProIleGlnSerValGlyLysProSerSerIleIleLysAsn 1140
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3496 .....ACCAAGAGAGCAAGTGTGTGAAGACGT 3524
1140 lThrThrLysProSerAlaThrAlaLysTyrThrGluLysGluSerGlu 1157
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3525 GGTG.....GACTCTAGTCAGA 3541
1157 lProGluLysLeuGlnLysLeuProLysGluAlaSerHisGluLeuMet 1173
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3542 AACCT..... 3546
1174 GlnHisGluLeuArgSerSerLysGlySerAlaSerSerGluLysGlyAr 1190
3546 ..... 3546
1190 gAlaLysAspArgGluHisSerGlySerGluLysAspAsnProAspLysA 1207
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3547 .....ACCCCATGCAAGAGAGATCTTCCCAA 3577
1207 rGlySerGlyAlaGlnProAspLysGluSerThrValAspArgLeuSer 1223
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3578 AGAAAGACAGTAGAGACCTCTCCAGAAAGCCGTCGAGAGAAAGAGT 3627
1224 GluGlnGlyHisPheLysThrLeuSerGlnSerSerLysGluThrArgTh 1240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3628 GAAGAGAGGAATCTCTCGGCCCTGAAATCAAAACGCGCACAC 3677
1240 rSerGluLysHisGluSerValArgGlySerSerAsnLysAspPhe... 1255
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3678 TCCAGCTTCAGAGAGTCAAGAAAGCAAGGTCTCCAGCAGCACTGGTCA 3727
1256 .....ThrProGlyArgAspLysLysValAspTyrAsp 1266
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3728 TCCCGCTCAGCACCTACTACAGAGACCGCAAGAAAA..... 3765
1267 SerArgAspTyrSerSerSerLysArgArgAspGluArgGlyGluLeuAl 1283
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3766 .....GAAGTCC 3773
1283 aArgArgLysAspSerProProArgGlyLysGluSerLeuSerGlyGln 1300
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3774 CAAGAACACTCTAGTGAGCCCAAGAAAGACAGCT..... 3810
1300 ySerLysLeuArgGluGluArgAspLeuProLysLysGlyAlaGluSer 1316
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3811 .....CCAGCAGCAATCAGTCCAGAGCAG 3837
1317 LysLysSerAsnSerSerProProArgAspLysLysProHisAspHisLy 1333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3838 AGCAAAACAGAAAAAGTGTGCTCCGCCCAAGTATCCCTGTAACAAADA 3887
1333 sAlaProTyrGluThrLysArgProCysGluGluThrLysProValAspL 1350
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3888 A...CCAAAGAGAAAGAAAGAACCA.....CCTCCGGTCATTA 3922
1350 yS..... 1350
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3923 AGCAGAGAAATGACAGCACTTTGAACATCTTCAGACATCTCTCCAAATGC 3972
1351 AsnSerGlyLysGluArg..... 1356
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3973 AATAGTCTTAAGCAAAAAATTCCAGACAGATGAGAGTCCAGAGATCAGAGT 4022
1357 ...GluLysHisAlaIleAlaIleArgAsnGlyLysGluSerSerGlyA 1372
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4023 GCACCTTAAGAGAGATGTGAAGCAAGAAATGTGTGGAGATGGAGGCT 4072
1372 lAsn..... 1373
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

4073 TAGGATCTTGACTCTGTCTTCTATACACCGAGGTTGCTTCTC 4122
1374CysHisVal..... 1376
4123 TGTGCAGTAGTGGCANGTAGAGTTTGTATTCACAGTCTGTGTGA 4172
1377TyrLeuThrArgLinhTrleupro.TTprArg 1387
4173 GCCCTCCACAAAGTTTGTATTAGAGAGACGAGCCCTCTGAGGACC 4222
1388 SerTrpLeuLeuGly 1392
4223 AGCTGGAATAATGCT 4237
seq_name: /cgn2_6/prodata/1/ina/5A_COMB.seq:US-08-306-691B-55
seq_documentation_block:
; Sequence 55, Application US/0830691B
; Patent No. 5734039
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Gonda, Lavoorgna & Monaco, P.C.
; STREET: Two Penn Center, Suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 KB
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306,691B
; FILING DATE: September 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEEX: No. 5734039e
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14255 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-306-691B-55
alignment_scores:
Quality: 355.00 Length: 1489
Ratio: 0.590 Gaps: 64
Percent Similarity: 40.430 Percent Identity: 19.946
alignment_block:
US-09-811-045A-1 x US-08-306-691B-55 ..
Align seg 1/1 to: US-08-306-691B-55 from: 1 to: 14255
90 LeuLeuGluSerAspLysHisThrCysProThrCysHisGlnAsnArg 106
..... |||:||||:|||||

892 ATTATCTCTTCAAAAAGACA..... 915
106 lSerProAspAlaLeuLleAlaAsnLysPheLeuArgGlnAlaValAsn 123
916GATGCACCAATGTGTAACCAACTTTACAGAGGACA..... 951
123 snPheLysAsnGluThrGlyTyrThrLysArgLeuArgLysGlnLeuPro 139
952AAAAAGGGGCTCAAAAGAAATTTGAAAAAGAACGACACT 990
140 ProPhe.....LeuPh 143
991 CAGCTGCAGAGAGAAAGGTGAACACAGCTCAAAATATTCAGACAGT 1040
143 eleuValProProProArgProLeuSerGlnArgAsnLeuGlnProArg 160
1041 CATCATGCGCTGTGTGACGTGCTATCTCCGCGGATCATTAAGACCCCTC 1090
160 erArgSerProLleLeuArgGlnGlnAsnProVal..... 172
1091 GCGGCTTTATAGAGATGAGATTTATGACCTTCATTAATAATTGCCCGA 1140
173PheArgTyrThrValSerProThrCysSerAs 183
1141 TTAGAGTCTACACCGAATAGTAGATTCACT...GCCCGCTCGTGAGATC 1187
183 pThrLysThrAlaGlySerCysSerAsp.....SerGlyT 195
1188 TTCTGAAAAATCAAGTCAGCTTCACGACCTCTCAATTAATGCTTCAG 1237
195 hrLeuSerArgLeuProAlaProSerLleSerSerLeuThrSerAsnGln 211
1238 ACTCTCTCGATCTAGTACGCCACGTGTGATACCTCCACAGACTCTCG 1287
212 SerSer.....LeuAlaProValSerGlyAsnProSe 223
1288 GCTTCTGAGAGATTCAGGTACTCTCTGAGAGCGGAGATACCCCTGA 1337
223 rSerAlaProAlaProValProAspLleThrAlaThrValSerLleServ 240
1338 AGTTCATCT...CCACTGCC.....ATTGCCACTGCC 1369
240 aLHisSerGluLysSerAspGlyProPheArgAspSerAsnLysLeu 256
1370 CAGAAATGAGATATGATAGACAGACAGAAAGATTCA..... 1410
257 LeuProAlaAlaAlaLeuThrSerGlnHisSerLysGlyAlaSerSerLl 273
1411CTGTCGAGAGACAGTTTGGATTCAGAACG... 1440
273 eaLalThrAlaLeuMetGluGlyGlyValProGlyThrSerProT 290
1440 1440
290 rPAsnSerLlePheValGlyGlnSerLeuHisGlyGlnLeuLeuPro 306
1440 1440
307 ThrThrGlyProValArgLleAsnAlaAlaArgProGlyGlyArgPr 323
1440 1440
323 oGlyTrrGluHisSerAsnLysLeuGlyTyrLeuValSerProProGlnG 340
1441ACGAAAAAATTATCAACTCTACAAAGTGCCTCCACAG. 1476
340 lnlleArgArgGlyGluArgSerCysTyrArgSerLleAsnArgGlyArg 356
1476 1476
357 HlSHlSerGluArgSerGlnArgThrGlnSerProSerLeuProAlaThr 373
1477CAGAGACCTCTCTGCTC..... 1494

```

373 rProCysPheValProValProProProLeuTyrProProProProH 390
      |||||
1495 .....CCACCTCCACCTCTCTGCTACCTCCACCGCCA. 1524
390 1sthrLeuProLeuProProGly..... 397
      |||||
1525 .....CCACTGCAGCCAGCTCCAGTATCTGACACACACCTTGG 1566
398 ..ValProPro.....ProGlnPheSerProGlnPheProSerSergl 411
      :|||
1567 CTTATGCTCTCAACATATCCCTTAGCATCACATTTTGGCTGCTTCCAC 1616
411 nProProThrAlaGlyTyrSer.....ValProProProGlyPhe. 424
      |||
1617 TGCCTCATGCAAGGGAAGGAAATATATTTCGAGAACCGCATTTA 1666
425 .....ProProAlaProAlaAsnIleSerThr 433
      |||
1667 GGTGACTCTTTTAAAGCATTTCTAGTCAAGCCACATATCTTTCCTCA 1716
434 AlaCysPheSer...ProGlyVal.....ProThrAlaHisSerAsnTh 447
      |||
1717 GCAAAAGTATGCCAAAGAGGCTTATTTCGCAAAACCAATATTGATTAATTT 1766
447 rMeProThrThrGlnAlaProLeuLeuSerArgGluIuPheTyrArg 464
      |||
1767 CCGACCCCTCCACCTACTCC..... 1788
464 lueGlnAsnAspIysGlyArgGluSerLysPheProTyrSergIysSer 480
      |||
1789 .....GAGGACGTTGGCTTGCATCTGTTTTTCTGCTGATCCGCT 1833
481 TyrSerArgSerSerTyrThrAspSerSergIuGlyLeuAlaGlnHisI 497
      |||
1834 GCTTCAGCCCGGATTTGTTTCGCCACTCATTCGGAACAAGGTTGATAT 1883
497 eHisAlaLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuH 514
      |||
1884 GCAC.....AAAGAGCCCT.....CTTCTGA 1906
514 1stAspHisProHisProProGluGluAlaGluAlaArgSerAlaMetIle 530
      :|||
1907 GAGCTCCAAAGATTACTCTCAAGTGAGGCTCCTAGATATTTGAGCTCT 1956
531 ValHisMetProAspLeuMetAspIleAlaHisAlaArgSerArgSerPr 547
      |||
1957 GTAACTTGGCTTAGTAATCGAACTTCTGCTGGAACATCTTTCAGGAGT 2006
547 oProTyrArgArgTyrArgSerArgSerArgSerProProGluPheArg 564
      |||
2007 ATCCAAATAGAAAAAGAAAGAAAGGTTTATGCTATTCGATCTGAC 2056
564 lueGlnSerProThrLysArgAsnValProArgGluGluLysGluArgGlu 580
      :|||
2057 CAAGATCTCTCTCT.....CACTCCATGAGACAAAGATGGAAGGCTT 2100
581 TyrPheAsnArgTyrArgGluValProProProTyrAspIleLysAlaTyr 597
      :|||
2101 AGTAGTCTAGCTCTACCTCTACCCCGCTCTCTCTCTCTCTC 2150
597 rTyrGlyArgSerValAsp.....P 604
      :|||
2151 GTTAAGCATTTCTGTAGTCTCTTTCGCACTAGTGCCTTAACCCACTT 2200
604 heArgAspProPheGluLysGluArgTyrArgGluTyrPgluArgLysTyr 620
      |||
2201 TTACTTTCTCTCT..... 2214
621 ArgGluTyrPgluLysTyrTyrLysGlyTyrAlaValAlaGluInPr 637
      :|||
2215 .....CATTCCTGACTCACTGCTGGGGAATCTGCAGAGAAAAATCA 2255

```

```

637 oArgPro.....SerAlaAsnArgGluAspPhe.....S 647
      |||||
2256 GAGACCAAGAGAGACAGACTAGTGTCCGGCAGAGCCATTTTCATCAAGTA 2305
647 eProGluAlaGluLeuProLeuAsnIleArgAsnSerProPheThrArg 663
      |||||
2306 GTCTACTCTCTCTCTCTCTGTTTACCCAGCTCTCAGACTGAAAGA 2355
664 GluYArgGluLysPyrAlaAlaGlyLysSerHisArgAsnArgAsnLe 680
      |||||
2356 GGGAGAAATAAAGCAAGGCC..... 2376
680 uGlyGlyAsnTyrProGluLysLeuSer...ThrArgAspSerHisAsnA 696
      |||||
2377 .....CCGAGAGCTGTCCAAAGATCGAGATGCTGCACAGA 2413
696 lAluYAspAsnProLysSerLysGlyLysGluSerGluAsnValProGly 712
      :|||
2414 GCGTGGAGAAAGGACMAAGATGAGAGAGAGACCGGAGAGA..... 2454
713 AspGlyLysGlyAsnLysHisLysLysHisArgLysArgAsnGluI 729
      |||
2455 ...GAAGAGAGATTAAGCGGAGCTCAAGAAAGAGAAAGAAAAAGG 2501
729 uLysGlyGluGluSerGluSerPheLeuAsnPro..... 740
      :|||
2502 ATCAGAAATTCAGAGTAGTCTGCTTGTATCCGTGGTAGGTTTCCA 2551
741 .....GluLeuGluGluThrSerArgLysCysArgGly 751
2552 MAGAAGAGTTGTTGGTGAAGATGTTGCCACTCATCTTTCGCAAAAAA 2601
752 SerSergIleAspGluThrLysThrAspThrLeuPheValLeuProse 768
      :|||
2602 GCACACAGCGCGAAGAGTCTCA.....TC 2627
768 rArgAspAspAlaThrProValArgAspGluPrometAspAlaGluSerI 785
      :|||
2628 ACATGATTTCTGGGACT.....GATATTACTCTCTG 2656
785 lethrPheLysSerValSerAspLysAspLysArgGluLysAspLysPro 801
      :|||
2657 TGACTCTT..... 2664
802 LysValLysSerAspLysThrLysArgLysSerAspLysSerAlaThrAl 818
      |||
2665 .....GGGATACACAGC 2678
818 aLysLysAspAsnValLeu...LysProSerLysGlyProGlnGluLysV 834
      |||
2679 TGTCAAACCAACCAATACTTATAAGAAAGGAGAGGAAATCTGGAAAAA 2728
834 alaSpGlyAspArgGluLysSerProArgSerGluProProLeuLys 850
      :|||
2729 CCAACTTGGAC.....CTGGGCCA...ACTGCCCATCTCTGGAGAAAG 2769
851 AlaLysGlu.....GluAlaThrLysIleAspSerValLysProse 864
      |||
2770 GAGAAACCCCTCTGCCCTTTCACACTCTTCATTCAGCACTGTAAACATTC 2819
864 rSerSerSer.....GlnL 869
      :|||
2820 CACTTCTCTCATAGGCTCCATGTTGGCTCAGACAGACAAAGCTTCCAAATGA 2869
869 yAspGluLysValThrGlyThrProArgLysAlaHisSerLysSerAla 885
      :|||
2870 CTGACAAAGAGGTTGCGAGCTCTTAATAAAGGCCAAAGCTCAGCTCTGC 2919
886 LysAspThrArgArgGluSerGlnProArgThrArgArgSerLysArgTh 902
      |||
2920 AAGATTGAGAAAGATAGAGTCTTAACCAACGACGACAG..... 2958
902 rValProLysThrSerSergIuLysSerGlnProValArgThr.....A 917

```



```

2959 .....CCCAAGCAGGTCACAGAAAGTGCATCATCAGACCTCTGTGC 3004
917 rgrtproragserleuargylsleasntryrleuilealaarglulys 933
3005 GAGAGACCCCGG.....ATTAAACAT...GTCTCAGAGAAGCA 3039
934 Asnqluarglulysargylserservalasplysasphegluserse 950
3040 GCGTTCCTCCCTGGCCGAAAACGAGCTGTCTTCCTCAT..... 3078
950 rsermetlysileservalgluylthrqluileallysproserp 967
3078 ..... 3078
967 rqlsargylsmetglulysaspvalglulysleuqluargthrpro... 982
3079 .....GACATGCCACCTGAGTGCCTTACCATG 3108
983 ...glulysaspylslealaserthrthrproalalsylsilely 998
3109 GAGAGACGAGAAAAGATTGTCTTC..... 3135
998 sleuasnaargluthrlyslsylegluasnaalagluasnaslaser 1015
3136 .ATGGGGAATGATACACAGTCATCAATTGCTGCTCAGAAATGCTGAAC 3184
1015 hrthrlysgluproserglulysleuqluser...thrserlysile 1030
3185 CTCTGTCTCCACCCATCAACCAATTAACTGTCACTAGAACCAAGCA 3224
1031 lysglnglulys.....Vallysgllysalalysarglusalagl 1045
3235 CCCAGGAACTCCAGTAAGAAAGCGTCGATCGAGCGGTGTGGCA 3284
1045 ysarglulys.....SerSerSerThrleuvalasplyr 1057
3285 GTGTCCCGCTGCCAGGTGCTGAGACTGTGTGTGTAATTCCT 3334
1057 hrserthrserthrlygllyserprovalargylserserglulys 1073
3335 TAGATAGCCCAAGTTGTGTGTGCGCAATATATAAGAACAGTGTGCAAG 3384
1074 Thrasphrlysarthrvalilelysthrmetgluluyrinasnas 1090
3385 .....ATGAGAAAATGTCGAATCT 3404
1090 psnthrAlaProAlaGluaspvalileilemetilegluvalproglins 1107
3405 ACATGATGCTCTCCAAAGCCTACCTGCAAGAACAGCAAACTGTGA 3454
1107 erlystrpasplysaspaspheglusergluqluasppvalylsThr 1123
3455 AAAAGAAAGAGAAAAGCTTAAGCACGTGAAAAGAAAGAC..... 3495
1124 ThrGlnProIleGlnserValGlylyserSerSerleilelysasna 1140
3496 .....AGCAAAAGAGAGCGAGTGTGTGAAGAACGT 3524
1140 lthrthrlyserSerAlaThrAlaLysTyrThrGluLysGluSerGlu 1157
3525 GGTG.....GACCTGTGTGACA 3541
1157 lnrproglulysleuqlulysleuProLysglulalaserHisgluLeu 1173
3542 AACCT..... 3546
1174 GlnHisGluLeuArgSerSerLysGlySerAlaSerSerGluLysGly 1190
3546 ..... 3546
1190 galalysasparglulhissergllyserglulysaspaspnproasplys 1207

```

```

3547 .....ACCCATCAGCAAGAGAGATCTGCCCAA 3577
1207 rqlserserglyalaglnproasplysluserthrvalaspargleuser 1223
3578 AGAAAGACAGTATGAGACTCTCCACGAAAGCCCGTCGAGAAAAGAGT 3627
1224 GluGlnGlyHisPheLysThrLeuSerGlnSerSerLysGluThrArg 1240
3628 GAAAGAGGAATGCTCGGCCCTGGGCTGATCCAAACAGGCCACAC 3677
1240 rserglulysHisGlnserValarglylserSerAsnLysasphe... 1255
3678 TCCAGCTTCCAGGAAGTCAAGCAAGAGTCTCCACAGCTGCTGCA 3727
1256 .....ThrProGlyArgaspLysLysValasPTyrAsp 1266
3728 TCCCGCTCAGCCACCTACTACAGGACGCCAAGAAAA..... 3765
1267 SerArgasPtyrSerSerSerLysArgArgaspLuarglulueal 1283
3766 .....GAGTTC 3773
1283 aargarglyaspsrProargllylsgluserleuserglul 1300
3774 CAAGAACACTCTCTAGTGCACCAAGAAAAGCAGCT..... 3810
1300 yserserleuarglulguarglueuprolyslsyalalaguser 1316
3811 .....CCACCACCAAGATCAGCTCCAGACAG 3837
1317 LysLysSerAsnSerSerProarGaspLysLysProHisPhtsly 1333
3838 AGCAAAAGAAAAAAGTGCTCCCGCCCAAGTCTGTAAACAAAA 3887
1333 salaprotlygluthrlysarproCysGluGluThrLysProValapl 1350
3888 A...CCAAAGAAAAGAAAACCA.....CTCCGGTCAATA 3922
1350 ys..... 1350
3923 AGCAGGAAATGCAGGACTTTGAACATCTCAGACACTCTCCAAATGC 3972
1351 AsnserGlyLysGluArg..... 1356
3973 AATGATTCTAAGCAAAAATTCACGAGATGAGACTCCACAGATCAAGT 4022
1357 ...GluLysHisAlaAlaGluAlaArgAsnGlyLysGluSerSerGly 1372
4023 GGACTTTAAGAGGATGTTGTAGAGAGAAAATGTGTGGAGATGGAGGCT 4072
1372 laasn..... 1373
4073 TAGGAATCTTGACTTCTGTCTTAACACCCAGGAGTGTGTCTTC 4122
1374 .....CysHisVal..... 1376
4123 TGTCACGATAGTGGCATGTAAGACTTTGTGTATGCCAAGTCTGTGTA 4172
1377 .....TyrLeuThrArgGlnThrLeuPro...TyrArgArg 1387
4173 GCCCTTCCACAAGTTTGTGTAGAGAGAACGAGCGCTGTGAGAGACC 4222
1388 SerThrPleuengly 1392
4223 AGCTGAAAATGCT 4237

```

```

seq_name: /cgn2_6/prodata/1/ina/6A_COMB.seq:us-08-545-860D-1
seq_documentation_block:
; Sequence 1, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo

```

APPLICANT: Canaan, Eli
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
 TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
 ADDRESSEE: No. 6040140r1s
 STREET: One Liberty Place, 46th floor
 City: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/545,860D
 FILING DATE: 07-MAR-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/04496
 FILING DATE: 22-APR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/10930
 FILING DATE: 09-DEC-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/327,392
 FILING DATE: 19-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/320,559
 FILING DATE: 11-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/062,443
 FILING DATE: 14-MAY-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/971,094
 FILING DATE: 30-OCT-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/888,839
 FILING DATE: 27-MAY-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/805,093
 FILING DATE: 11-DEC-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Deluca Esq., Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: TJU-1262
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 14255
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 ANTI-SENSE: NO
 US-08-545-860D-1


```

902 rValProLysThrSerSerGlnLysSerGlnProValArgThr.....A 917
    |||||.....|
2959 ...CCCAAGACAGAGGCTCAAGAAAGTCACTCATCAGAGACCTCTGTCG 3004
    |||||.....|
917 rGArgProArgSerLeuArgLysIleAsnTyrLeuIleAlaArgGluLys 933
    |||||.....|
3005 GAGGACCCCGG.....ATTAAACAT...GTCTGCAGAGAGCA 3039
    |||||.....|
934 AsnGluArgGluLysArgLysSerValAspLysAspPheGluSer 950
    |||||.....|
3040 GCTGTGGCCCTTGCCGGAACGAGCTGTGTTCTCAT..... 3078
    |||||.....|
950 rSerMetLysIleSerLysValGluGlyThrGluIleValLysProSer 967
    |||||.....|
3078 ..... 3078
967 rGlyArgLysMetGluGlyAspValGluLysLeuGluArgThrPro... 982
    |||||.....|
3079 .....GACATGCCACCCCTGAGTGCCTTACCATTGG 3108
    |||||.....|
983 ...GluLysAspLysIleAlaSerSerThrThrProAlaLysLysIle 998
    |||||.....|
3109 GAGGAACGAGAAAGATTGTTGTTCTTC..... 3135
    |||||.....|
998 rLeuAsnArgGluThrGlyLysLysIleGlyAsnAlaGluAsnAlaSer 1015
    |||||.....|
3136 .ATGGGGAATGATGACAGTCAATGCTGCTGCAGAGATGCTGAAAC 3184
    |||||.....|
1015 hrThrLysGluProSerGluLysLeuGluSer...ThrSerLysIle 1030
    |||||.....|
3185 CTCTGCTCCACCCATCAACCAATTAAACCTGCTCAGAAACAGCA 3234
    |||||.....|
1031 LysGlnGluLys.....ValLysGlyLysAlaLysArgLysValAlaGlu 1045
    |||||.....|
3235 CCCCAAGAACCTCCAGTAAGAAAGACGCTCATGAGCGGTGGGCA 3284
    |||||.....|
1045 ySerGluGly.....SerSerThrLeuValAspTyrT 1057
    |||||.....|
3285 GTGTCCCGGCTGCAGGTCCTGAGAGCTGTGTGTTGTTACTAATGCT 3334
    |||||.....|
1057 hrSerThrSerThrGlyGlySerProValArgLysSerGluLys 1073
    |||||.....|
3335 TAGAATAGCCCAAGTTGTGTGCTCCAAATATTAAGAACAGTGTCTG 3384
    |||||.....|
1074 ThrAspThrLysArgThrValIleLysThrMetGluGluTyrAsnAsn 1090
    |||||.....|
3385 .....ATGAGAAATGTCAGAACT 3404
    |||||.....|
1090 pAsnThrAlaProAlaGluAspValIleIleMetIleGlnValProGln 1107
    |||||.....|
3405 ACAATGATGCTCTCCAAAGCTTACCTGCAGAGACAGCTAAAGCTGCA 3454
    |||||.....|
1107 eLysTrpAspLysAspAspPheGluSerGluGluGluAspValLysThr 1123
    |||||.....|
3455 AAAAAGAAAGCAAAAGTCTAAGACAGTGAAGAAAGAAC..... 3495
    |||||.....|
1124 ThrGlnProIleGlnSerValGlyLysProSerIleIleLysAsnVal 1140
    |||||.....|
3496 .....AGCAAGAGACAGTGTGTGAAGAACCT 3524
    |||||.....|
1140 lThrThrLysProSerAlaThrAlaLysTyrThrGluLysGluSerGlu 1157
    |||||.....|
3525 GGTG.....GACTCTATTCAGA 3541
    |||||.....|
1157 lnrProGluLysLeuGlnLysLeuProLysGluAlaSerHisGluLeuMet 1173
    |||||.....|
3542 AACCT..... 3546
    |||||.....|
1174 GlnHisGluLeuArgSerSerLysGlySerAlaSerSerGluLysGlyAr 1190
    |||||.....|
3546 ..... 3546

```

```

1190 gAlaLysAspArgGluHisSerGlySerGluLysAspAsnProAspLysA 1207
    |||||.....|
3547 .....ACCCATTCAGACAGAGAGATCTCTGCCCA 3577
    |||||.....|
1207 rGlySerGlyAlaGlnProAspLysGluSerThrValAlaAspArgLeuSer 1223
    |||||.....|
3578 AGAAAGCAGTAGTAGGCTCTCCACGAAACCCGTCGAGAAAGAGT 3627
    |||||.....|
1224 GluGlnGlyHisPheLysThrLeuSerGlnSerSerLysGluThrArgThr 1240
    |||||.....|
3628 GAGAAAGGAATGCTCGGCCCTGAGCTGATCCAAACAGACCCACAC 3677
    |||||.....|
1240 rSerGluLysHisGluSerValArgGlySerSerAsnLysAspPhe... 1255
    |||||.....|
3678 TCCAGCTTCCAGAGAGTCAAGACAGAGGTCCTCCAGCAGACGTGCA 3727
    |||||.....|
1256 .....ThrProGlyArgAspLysValAspTyrAsp 1266
    |||||.....|
3728 TCCCGCTCAGCCACCTACTACAGAGACCCCAAGAAA..... 3765
    |||||.....|
1267 SerArgAspTyrSerSerSerLysArgArgAspGluArgGlyGluLeuAl 1283
    |||||.....|
3766 .....GAACTTC 3773
    |||||.....|
1283 aArgArgLysAspSerProProArgGlyLysGluSerLeuSerGlyGlnL 1300
    |||||.....|
3774 CAAAACCACTCCTAGTGAAGCCCAAGAAAGCAGCCT..... 3810
    |||||.....|
1300 ySerLysLeuArgGluGluArgAspLeuProLysLysGlyAlaGluSer 1316
    |||||.....|
3811 .....CCACCACCAAGTACAGTCCACAGAG 3837
    |||||.....|
1317 LysLysSerAsnSerSerProProArgAspLysLysProHisAspHisL 1333
    |||||.....|
3838 AGCAAAACAGAAAGATGCTGCCGCCCAAGTATCCCTGTAACAAAGAA 3887
    |||||.....|
1333 sAlaProTyrGluThrLysArgProCysGluGluThrLysProValAspL 1350
    |||||.....|
3888 A...CCAAAGAAAGAAAGAAACCA.....CCTCGGTCAATA 3922
    |||||.....|
1350 yS..... 1350
    |||||.....|
3923 AGCAGAGAAATGCAGCACTTGAACATCTCAGCACTCTCCAATGCG 3972
    |||||.....|
1351 AsnSerGlyLysGluArg..... 1356
    |||||.....|
3973 AATATTTCTAAGCAAAATTCACAGACATGAGTCCACAGATCAGAGT 4022
    |||||.....|
1357 ...GluLysHisAlaAlaGluAlaArgAsnGlyLysGluSerSerGlyA 1372
    |||||.....|
4023 GGACTTTAAGAGAGATGTGAAGACAGAAATGTGTGGAGATGGAGGCT 4072
    |||||.....|
1372 lAsn..... 1373
    |||||.....|
4073 TAGAATCTTGACTTCTGTTCTATTAACACCCAGGGTGTGCTTCTTC 4122
    |||||.....|
1374 .....CysHisVal..... 1376
    |||||.....|
4123 TGTGCCAGTAGTGGCAGTGAAGTGTGTATTTCCCAAGTCTGTGGCA 4172
    |||||.....|
1377 .....TyrLeuThrArgGlnThrLeuPro.TyrArgArg 1387
    |||||.....|
4173 GCCCTTCCACAAGTTTGTGTTAAGAGAGAAAGAGCGCCTGTGGAGACC 4222
    |||||.....|
1388 SerTrpLeuLeuGly 1392
    |||||.....|
4223 AGCTGAAATAATGGT 4237
    |||||.....|
seq_name: /cgn2_6/ptoddata/1/ina/PCTUS_COMB.seq.PCT-US94-04496-1
seq_documentation_block:
; Sequence 1, Application PC/TUS9404496
; GENERAL INFORMATION:

```

APPLICANT: Croce, Carlo
 APPLICANT: Cnaan, Eli
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
 TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1
 NUMBER OF SEQUENCES: 86
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
 ADDRESSEE: Norris
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/04496
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Deluca Esq., Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: TJD-1242
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 14255
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 ANTI-SENSE: NO

```
alignment_scores:
  Quality: 355.00
  Ratio: 0.590
  Percent Similarity: 40.430
  Length: 1489
  Gaps: 64
  Percent Identity: 19.946
```

```
alignment_block:
US-09-811-045A-1 x PCT-US94-04496-1    ..
```

Align seg 1/1 to: PCT-US94-04496-1 from: 1 to: 14255

```

90 LeuLeuGluSerAspAspLysHisHisThrGlyProThrGlySerHisGlnAsnAspVal 106
    ::::: ||||| ::||| ::||| ::|||
892 ATTATTCCCTCTTCAAAAAGACA..... 915

106 IserProAspAlaLeuIleAlaAsnLysPheLeuArgGlnAlaValAsnA 123
    |||||
916 .....GATGACACCATTTGCTTAAACCAACTCTTACAGAGGCA..... 951

123 snPheLysAsnGlnIuThrGlyTyrThrLysAspGluArgLysGlnLeuPro 139
    :: ||| ||||| ::||| ::|||
952 .....AAAAAGGGCGCTCAAAAGAAATTTGAAAAAAGAACCACT 990

140 ProPhe.....LeuPh 143
    ::::

991 CAGCTGCGACGGAAGAAGGTGAAGACACAGGTCAAAAATATTTCGACATT 1040

143 LeuValProProArgProLeuSerGlnArgAsnLeuProArgS 160
    |:::||||| ::||| ::||| ::|||
1041 CATCATGCTGTTGTTCAGATGCTATCTCCTCGGGATCATTTAAGACCCCTC 1090

160 eArgSerProIleLeuArgGlnGlnAspProValVal..... 172
    ||| ::||| ::||| ::|||
1091 GGCCTGTTATAGAGATGAGGATTTTGACCCCTCCCAATTAAAATTCGCCGA 1140

```

```

173 ..... PheArgThrAlaSerProHisSerAsp.....SerGly 195
1141 TTAGAGCTACACCGAATGATGATTCAGT...GCCCGCTCTGTGGATC 1187
183 rPhrArgThrAlaGlySerCysSerAsp.....SerGly 195
1188 TTTCTGAAAAATGACATGGACGTTCTTCACACGCTCTCTCAATATGTTCTCAG 1237
195 hLeuSerArgLeuProAlaProSerIleSerSerLeuThrSerAsnGln 211
1238 ACTCGCTCTGCACTAGTACGACCCGACATGTGTGATCTCCACAGACTCTCAG 1287
212 SerSer.....LeuAlaProProAlaSerGlyAsnPro 223
1288 GCTTCTGACGACATTCACAGTACTCTCTGACGACGACGACGATACCCCTGCA 1337
223 rSerAlaProAlaProAlaProAspIleThrAlaThrValSerIleSerV 240
1338 ACTGTCATCT...CCACGAGCCC.....ATTGCCACAGTCCC 1369
240 aHisSerGlyLeuSerAspGlyProPheArgAspSerAspAsnLeu 256
1370 CAGAAATACAGTATGATGATACGACACACGACGATATCA..... 1410
257 LeuProAlaAlaAlaLeuThrSerGlnHisSerGlyAlaSerSerI 273
1411 .....GTGCGGAGAGAAAGTTTGGATCTAGACAG... 1440
273 eAlaIleThrAlaLeuMetGluGluGlyValProGlyThrSerProT 290
1440 ..... 1440
290 rAsnSerIlePheValGlyInsSerLeuHisGlyGlnLeuIlePro 306
1440 ..... 1440
307 ThrThrGlyProValArgIleAsnAlaAlaArgProGlyGlyAlaArg 323
1440 ..... 1440
323 oGlyThrGlnHisSerAsnLeuGlyThrLeuValSerProProGln 340
1441 .....ACGAAAAATTTCACACGCTACAAAGTCCGCCCCGAC 1476
340 hIleArgArgGlyLeuArgSerCysTyrArgSerIleAsnArgGlyArg 356
1476 ..... 1476
357 hHisSerGlyArgSerGlnArgThrGlnSerProSerLeuProAlaThr 373
1477 .....CAGGACGCTCTCTGCT..... 1494
373 rProCysPheValProValProProProLeuThrProProProGln 390
1495 .....CCACGCTCACGCTGCTGACGCTCCACCGCCA 1524
390 iThrLeuProLeuProProGly..... 397
1525 .....CCACTGCACGACGCTCCACTATCTGACACACACGCTTGG 1566
398 ...ValProPro.....ProGlnHisSerProGlnPheProSerSerG 411
1567 CTATATGCTCTCCACAAATCCCTTACGATACACATTTTGGCTGCTCCAC 1616
411 nProThrAlaGlyTyrSer.....ValProProProGlyPhe 424
1617 TGTCTCTATGCAAGGAGACGAAATATGATTTCGCGAAGACGACATTTA 1666
425 .....ProProAlaProAlaAsnIleSerThr 433
1667 GGTGACCTCTTTAAAGCATCTCTAGTGCAGGACGCAATATCTGTTCTGCA 1716

```

```

434 AlaCysPheSer...ProGluVal.....ProThrAlaHisSerAsn 447
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1717 GCMAAGTATGCCAAGAGAGGTCTTATTCGCAACCAATATTTGATATTT 1766
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
447 rMetProThrThrGlnAlaProLeuLeuSerArgGluPheTyrArgS 464
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1767 CCGACCCCTCCACTACTCC..... 1788
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
464 LuGlnAsnAspLysGluArgGluSerLysPheProTyrSerGlySerS 480
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1789 ....GAGACGTTGGCTTGCATCGGTGTTTCTGCATCTGCTACCGCT 1833
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
481 TyrSerArgSerSerTyrThrAspSerSerGlnGlyLeuAlaGlnHis 497
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1834 GCTTCAGCCCGATGTTTTCGCCACTCCATCTCGAACAAGGTTTGATAT 1883
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
497 eHisAlaLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHis 514
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1884 GCAC.....AAAAGAGCCCT.....CTTCTGA 1906
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
514 IsAspHisProHisProGluGluAlaGluAlaArgSerAlaMetIle 530
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1907 GAGCTCCAAATTTACTCCAAAGTAGGCTCATCTAGATATTTGAGTCT 1956
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
531 ValHisMetProAspLeuMetAspIleAlaHisAlaArgSerArgSerP 547
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1957 GTAACTCTGCTAGTATCGAACTTCTCGAACAATCTTCTTCAGAGT 2006
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
547 oProTyrArgArgTyrArgSerArgSerArgSerProGluPheArgS 564
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2007 ATCCAAATGAAAAAGAAAAAGAAAGTTTATGCTATTTCGATCTGAAC 2056
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
564 LysGlnSerProThrLysArgAsnValProArgGluGluLysGluArgGlu 580
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2057 CAAGATCTCTCTCT....CACTCCATGAGGACAGAAAGTGAAGGCTT 2100
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
581 TyrPheAsnArgTyrArgGluValProProProTyrAspIleLysAlaTy 597
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2101 AGTAGTCTGAGCTCTCACCTCACCCTCCACCCCGTCTCTCTCTCTC 2150
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
597 rTyrGluArgSerValAsp.....P 604
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2151 GTTAAGCATTTCTGTAGTCTCTTGCACATAGCTTAACCCACTT 2200
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
604 heArgAspProPheGluLysGluArgTyrArgGluTyrPheGluArgLysTyr 620
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2201 TTACTTTCTCTCT..... 2214
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
621 ArgGluTyrTyrGluLysTyrTyrLysGlyTyrAlaValGlyAlaGlnP 637
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2215 .....CAATCCCTGACTCAGTCTGGGAATCTGCAGAAAAAATCA 2255
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
637 oArgPro.....SerAlaAsnArgGluAspPhe.....S 647
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2256 GAGACCAAGAGACAGACTAGTGTCCGCGACAGAGCCATTTCATCAATA 2305
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
647 eProGluArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArg 663
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2306 GTCCCTACTCTCTCTCTGCTTACCCCAAGCTCTCAGACTCAAAAGA 2355
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
664 GluArgArgGluAspTyrAlaAlaGlyGlnSerHisArgAsnArgAsnLe 680
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2356 GCGAGAAATTAAGACAAAGCC..... 2376
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
680 uGlyGlyAsnTyrProGluLysLeuSer...ThrArgAspSerHisAsnA 696
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2377 .....CCGAGAGAGCTGTCCAAAGATGAGATGCTGCACAAAGA 2413
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
696 IalysAspAsnProLysSerLysGluLysGluSerGluAsnValProGly 712
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2414 GCGTGGAGAGACAAAGATAGAGAGAGACGGGAGAGA..... 2454
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
713 AspGlyLysGlyAsnLysHisLysHisArgLysArgArgAsnGlu 729

```

```

2455 ....GAAAAGAGAAATAAGCGGAGTCAAGAGAAAGAAAAAGGG 2501
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
729 uLysGlyGluGlnSerGlnSerPheLeuAsnPro..... 740
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2502 ATCAGAAATTCAGATGATGTTCTGTTGATCTGTGGGTAGGTTTCCA 2551
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
741 .....GluLeuLeuGluThrSerArgLysCysArgGly 751
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2552 AAGAGAGAGTTGTTGGTGAAGATGTTGCCACTTCATCTTCTGCCAAAAA 2601
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
752 SerSerGlyTLeaspGluThrLysThrAspThrLeuPheValLeuProse 768
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2602 GCACAAGGGCGGAAGAAGCTTCA.....TC 2627
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
768 rArgAspAspAlaThrProValArgAspGluPrometAspAlaGluSerI 785
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2628 ACATGATTCYGGACT.....GATATTACTTCTG 2656
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
785 LeThrPheLysSerValSerAspLysAspLysArgGluLysAspLysPro 801
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2657 TGACTCTT..... 2664
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
802 LysValLysSerAspLysThrLysArgLysSerAspGlySerAlaThrAl 818
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2665 .....GGGATACACAGC 2678
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
818 aLysLysAspAsnValLeu...LysProSerLysGlyProGluGluLysV 834
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2679 TGICAAAAACCAAAATACTTAAAGAAAGGAGAGAAATCTGGAAGAAA 2728
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
834 aLysGlyAspArgGluLysSerProArgSerGluProProLeuLysLys 850
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2729 CCAACTGTGAC....CTGGGCCA...ACTGCCATCCCTGGAGAAG 2769
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
851 AlalysGlu.....GluAlaThrLysTLeaspSerValLysProse 864
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2770 GAGAAACCCCTCTCTTCACCTTCATCTGACCTGTAACATTC 2819
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
864 rSerSerSer.....GlnL 869
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2820 CACTTCTCCATAGCTCCATGTGGCTCAGGACAGACACATTCACATGA 2869
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
869 ysAspGluLysValThrGlyThrProArgLysAlaHisSerLysSerAla 885
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2870 CTGACAAAGGGTTCAGCTCTCTTAAGGCAAGGCAAGCTCAGCTCTGC 2919
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
886 LysAspThrArgArgGlnSerGlnProArgThrArgArgSerLysArgTh 902
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2920 AAGATTGAGAGAGTAAGAGTCTTAACAAACGACAGC..... 2958
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
902 rValProLysThrSerSerGlnLysSerGlnProValArgThr.....A 917
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2959 ....CCCAAGCACAGGCTCAAGAAAGTCACTCATCAGAGCTCTGTGC 3004
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
917 rGArgProArgSerLeuArgLysIleAsnTyrLeuIleAlaArgGluLys 933
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3005 GAGGACCCCGG.....ATTAACAT...GTCTGAGAGAGACA 3039
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
934 AsnGluArgGluLysArgLysLysSerValAspLysAspPheGluSerSe 950
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3040 GCTGTGGCCCTGGCGCAAAAGAGACTGTGTTCCCTGAT..... 3078
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
950 rSerMetLysIleSerLysValGluGlyThrGluIleValLysProserP 967
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3078 ..... 3078
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
967 rOLysArgLysMetGluGlyAspValGluLysLeuGluArgThrPro... 982
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3079 .....GACATGCCACCTTAGAGGCTTACCATAGG 3108
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
983 ...GluLysAspLysIleAlaSerSerThrThrProAlaLysLysIleLys 998
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```



```

: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Reiter, Stephen E.
: REGISTRATION NUMBER: 31,192
: REFERENCE/DOCKET NUMBER: P41 9387
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619)546-4737
: TELEFAX: (619)546-9392
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 11907 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: US-08-061-376-4

alignment_scores:
  Quality: 345.50      Length: 1568
  Ratio: 0.537         Gaps: 72
  Percent Similarity: 41.008   Percent Identity: 20.026

alignment_block:
  us-09-811-045a-1 x US-08-061-376-4  ..

Align seg 1/1  to: US-08-061-376-4  from: 1  to: 11907

18  GLYLSYRAlaIleProthrIleAspAlaGlu.AlaTyraIleGlyL 34
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
876  GGAAGAGAGGGGGTCAAAATTGTAGACGGAGAGAGAGGCCCTCCATCAACA 925
   34  ySLySLyLysProProPhe..LeuPro.....GluGluP 45
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
926  GAAAGATTAAGACCCCTTCGCTCCATATTCTGAACTCGAAGAAC 975
   45  rSeSerSerSerGluGluAspAspProIleProAlaGluLeuLeuCys 61
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
976  CCCAGAAAGTCGGAGAACACAGAGAACACCTCA..... 1013
   62  LeuIleCysLysAspIleMetThrAspAlaValIleProCysGSL 78
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1014  ..CTTACAAAGAA.....GATAGACAGTTGTC..... 1040
   78  yAsnSerSerCysAspGluCysIleArgThrThrLeuLeuGluSerAspL 95
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1041  .AGACAAAGCCCTCGAGAGATTAAAGCCAGTTAGATTATCTCTTCA 1089
   95  ySHSLThCysProThrCysSHSLGlnAsnAspValSerProAspAlaLeu 111
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1090  AAAGGACA.....GATCCAAC 1106
   112  IleIleAsnLysPheLeuArgIleAlaValAsnAsnPheLysAsnGluTh 128
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1107  ATTGCTAAGCAACTCTTACAGAGGCA.....AAAA 1138
   128  rGLyTyThr..LysArgLeuArgLysGlnLeuProProPhe..... 141
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1139  AGGGGGCTCAAAAGAAATTTGAAAAAGACGCTCAGCTCGACGGAGGA 1188
   142  .....LeuPheLeuValProProP 148
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1189  AAGGTGAAGACACAGGTCAAAAATATTCGACAGATCATCATGCTGTGT 1238
   148  cArgProLeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleL 165
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1239  CAGTCTATCTCTCTCGCGATCATTAAGACCCCTCGCGGTTTATAGAG 1288
   165  euArgLingLinspProValVal..... 172
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1289  ATGAGATTATGACCCCTCAATTAATAATGCCCGATTAGACTCAACCG 1338
   173  ..PheArgTyThrIleValSerProThrCysSerAspThrLysThrAlaG 188
```

```

   |||:|||||  |||:|||||  |||:|||||  |||:|||||
1339  AATAGTACATTCAGT...GCCCGTCTGTGGATCTTGAAAAATCAAG 1385
   168  ySerCysSerAsp.....SerGlyThrIleuSerArgLeuP 200
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1386  TGCACCTTCTCAGCACCTCTCAAAATGTTCACAGCTCTCTCATCTA 1435
   200  rAlaProSerIleSerSerLeuThrSerAsnGlnSer..... 213
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1436  GTAGCCCGAGTGTATATACCTCCACAGACTCTCAGAGCTTCTAGAGATT 1485
   214  ....LeuAlaProProValSerGlyAsnProSerSerAlaProAlaP 228
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1486  CAGTACTTCTCGAGACGCGAGCATACCCCTGAAGTTCAATCT...CC 1532
   228  oValProAspIleThrAlaThrValSerIleSerValHisSerGlyLys 245
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1533  ACTGCC.....ATTCCAGTCCCGACAGAAATGAGAGTA 1567
   245  eAspSGlyProPheArgAspSerAspAsnLysLeuLeuProAlaAla 261
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1568  ATGATAGAGAGAGACAGAGGTATTCA..... 1593
   262  LeuThrSerGluHisSerLysGlyAlaSerSerIleAlaIleThrAla 278
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1594  ..GTTCGAGAGAGAAATTGTGATCTAGAACG..... 1623
   278  uMetGluLysGlyValProGlyThrSerProThrAsnSerIlePheV 295
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1623  ..... 1623
   295  aLGIyLInSerLeuLeuHisGlyLInLeuIleProThrGlyProVal 311
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1623  ..... 1623
   312  ArgIleAsnAlaAlaArgProGlyGlyValArgProGlyTyrGlnHis 328
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1624  .....AC 1625
   328  rAsnLysLeuGlyTyLeuValSerProProGlnInIleArgArgLys 345
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1626  GAAAAATTTATCACTTACAAAGTGCCCCCAGCAG..... 1662
   345  IuArgSerCysTyArgSerIleAsnArgGlyArgHisHisSerGluArg 361
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1662  ..... 1662
   362  SerGlnArgThrInSerProSerLeuProAlaThrProCysPheValPr 378
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1663  ..CAGACCTCTCTGCTCA..... 1680
   378  oValProProProLeuTyTyProProPro..... 389
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1681  ...CTCCACCTCTGCTGACTCCACGCCACACACTGACGACGCTCA 1726
   390  .....HisThrLeuPro.....LeuProProGlyValProPro 400
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1727  GTATCTGTGACACACA...CTTGGCTTATGCCCAACATC..... 1767
   401  ProGlnPheSerProGlnPheProSerSerGlnProProThrAlaGlyTy 417
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1768  CCTTAGCATCACCACTTTTGTGCTGCTTCCACTGCTCTATAGCAAGGAA 1817
   417  rSer.....ValProProProGlyPhe..... 424
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1818  GCGAAAAATCTATTTTGGAGAAACGACATTTAGGAGGACTTTTAAAGC 1867
   425  ....ProProAlaProAlaAsnIleSerThrAlaCysPheSer...Pro 438
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
1868  ATTCTAGTCAAGAGCACAATATCTTCTCCACAGCAAAATATGCAAGAA 1917
   439  GlyVal.....ProThrAlaHisSerAsnThrMetProThrInGlnAl 453
       |||:|||||  |||:|||||  |||:|||||  |||:|||||
```



```

1918 GGTCTATTGCAAAACAATATTGATTTCCGACCCCTCCACTAAC 1967
453 aProleuLeuSerArgGluGluPheTyrArgGluGlnAsnAspLysGlyA 470
1968 TCCC.....GAGGACGTTGGCT 1984
470 rGluSerLysPheProTyrSerGlySerSerTyrSerArgSerSerTyr 486
1985 TTGCATCTGCTTTTCTGCATCTGTGACCGCTGCTTACCCCGATGTTT 2034
487 ThrAspSerSerGlnGlyLeuAlaGlnHisIleHisAlaLeuThrLeuSe 503
2035 TCGCCACTTCATCTGGACAAGGTTTGATATGCAC.....AAAGAG 2078
503 rProSerAlaAlaHisThrLeuAspLeuLeuHisAspHisProHisPro 520
2079 CCTT.....CTTCTGAGAGCTCCACAAGATTATCTC 2107
520 rGluGluAlaGluAlaArgSerAlaMetIleValHisMetProAspLeu 536
2108 CAAGTCAGGCTCACTCTAGATAATTGAGCTGTACCTTGCTTACTAAT 2157
537 MetAspIleAlaHisAlaArgSerArgSerProProTyrArgTyrArg 553
2158 CGAACTTCTGTGGAACATCTTCTTCAGAGATATCCAAATAGAAAAAGAA 2207
553 gSerArgSerArgSerProProGluPheArgGlyGlnSerProThrLysA 570
2208 AAGAAAAGTTTATGCTTATTCGATCTGACCAACAGATCTCTTCT..... 2253
570 rGAsnValProArgGluGluGluGluArgGluTyrPheAsnArgTyrArg 586
2254 ..CACCTCCATGAGACAGACAGAGGAGGCTTAGTGTCTGAGCTCTCA 2301
587 GluValProProProTyrAspIleLysAlaTyrTyrGlyArgSerValAs 603
2302 CCTTCACCCCGCCGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 2351
603 P.....PheArgAspProPheGlu 610
2352 TCCCTTGGCCACTAGTGCCTTAAACCACTTTTCTCTCTCT..... 2397
610 yGluLysArgTyrArgGluTyrGluArgLysTyrArgGluTyrPyrGluLys 626
2398 .....CATTCCTCG 2406
627 TyrTyrLysGlyTyrAlaValGlyAlaGlnProArgPro..... 639
2407 ACTCACTCTGGGGAATCTGCAGAGAAAAATCAAGACCAAGAGAGACAG 2456
640 .SerAlaAsnArgGluAspPhe.....SerProGluArgLeuLeuP 653
2457 TAGTGTCCGCGGAGCCATTTTCATCAAGTAGTCTTCTCTCTCTCTCC 2506
653 rLeuLeuAsnIleArgAsnSerProPheThrArgLysArgArgGluAspTyr 669
2507 CTGTGCTTACCCAGGCTCTCAGACTGAAAGAGGAGAAATTAAGACAA 2556
670 AlaAlaGlyIleSerHisArgAsnArgAsnLeuGlyLysAntyrProG 686
2557 GCC.....CCCGA 2564
686 uLysLeuSer...ThrArgAspSerHisAsnAlaLysAspAsnProLys 702
2565 GGGAGCTGTCCAAAGATCGATGCTGACAAAGACGTCGAGAGAGAGACA 2614
702 rLysGluLysGluSerGluAsnValProGlyAspGlySerGlyAsnLys 718
2615 GTAGAGAGAGACCGGAGAGAG.....CAAAAGGAGAAATTAAG 2652
719 HisLysLysHisArgLysArgArgAsnGluGluLysGlyGluGluSerG 735
2653 CGGAGTCAAGAAAGAAAGAAAGGAGATCAAGAAATTCAGAGTAG 2702

```

```

735 uSerPheLeuAsnPro.....G 741
2703 TTCTGCTTTGTATCTGTGGTAGGCTTCCAAAGAGGTTGTGGTG 2752
741 IuLeuLeuGluThrSerArgLysCysArgGlySerSerGlyIleAspGlu 757
2753 AAGATTTGCACTTCTCATCTTCCCAAAAAAGCAAGAGCGGAGAGAG 2802
758 ThrLysThrAspThrLeuPheValLeuProSerArgAspAspAlaThrP 774
2803 TCTTCA.....TCACATGATCTGCGGACT... 2826
774 oValArgAspGluProMetAspAlaGluSerIleThrPheLysSerValS 791
2827 .....GATATTACTTCTGTGACTTT..... 2847
791 eAspLysAspLysArgGluLysAspLysProLysValLysSerAspLys 807
2847 ..... 2847
808 ThrLysArgLysSerAspGlySerAlaThrAlaLysLysAspAsnValIe 824
2848 .....GGGATACACACAGCTGTCAAAAACCAAAATACT 2879
824 u..LysProSerLysGlyProGlnGluLysValAspGlyAspArgGlu 840
2880 TATTAAGAAAGGAGAGAAATCTGAAAAAACCACTTGAC.....C 2923
840 ySerProArgSerGluProProLeuLysLysAlaLysGlu..... 853
2924 TCGGCCCA..ACTGCCCATCCCTGAGAGAGAGAAAAACCTTGCCTT 2970
854 GluAlaThrLysIleAspSerValLysProSerSerSer..... 867
2971 TCCACCTTCATCTGACACTGTTAAACATTTCCACTTCTCCATAGAGCTC 3020
868 .....GlnLysAspGluLysValThrG 875
3021 CATGTGGCTCAGGACAGACAGCTTCCAAAGTGTGACAAAGGGGTGCCA 3070
875 LysThrProArgLysAlaHisSerLysSerAlaLysAspThrArgArgGln 891
3071 GCCTCTTAAAAAGGCCAAAGCTCAGCTGTGCAAGATGTGAAGAGTAAG 3120
892 SerGlnProArgThrArgArgSerLysArgThrValProLysThrSerSe 908
3121 AGTCTTAAACAAACGACAG.....CCCAAGCACAGGG 3155
908 rGlnLysSerGlnProValAlaThr.....ArgArgProArgSerLeuA 923
3156 TCAAGAAAGAGACTCATCAGAGACTCTGTGCGAGAGACCCCGG..... 3198
923 rGlyLysIleAsnTyrLeuIleAlaArgGluLysAsnGluArgGluLysArg 939
3199 .....ATTAAACAT...GTCTGCAGAAAGACACTGTGCTTGGCCGA 3240
940 LysLysSerValAspLysAspPheGluSerSerMetLysIleSerLys 956
3241 AAACGAGCTGTGTTCCTGAT..... 3261
956 sValGluGlyThrGluIleValLysProSerProLysArgLysMetGluG 973
3261 ..... 3261
973 LysAspValGluLysLeuGluArgThrPro.....GluLysAspLysIle 987
3262 ..GACATGCCCACTGAGTGTCTTACCATGGGAAGACAGAAAGAT 3309
988 AlaSerSerThrThrProAlaLysLysIleLysLeuAsnArgGluThrG 1004
3310 TTGCTCTCC.....ATGGGAAATGATGACAA 3335

```

```

1004 yLysIleGlyAsnAlaGluAsnAlaSerThrThrIysGluProSerg 1021
      ::::::::::::::::::::
3336 GTCATCAATGCTGCTCAGAGATGCTCAACCTCTGCTCCACCATCA 3385
      ::::::::::::::::::::
1021 LuIysLeuGluSer...ThrSerIysIleIysGluIlys... 1034
      ::::::::::::::::::::
3366 AACCAATTAAACCTGCTCACTAGAACAGCCACGACCTCCACATA 3435
      ::::::::::::::::::::
1035 ValIysGlyIysAlaIysArgIysValAlaGlySerGluGly... 1048
      ::::::::::::::::::::
3436 AAGAAAGACGCGCATCGAGCGGCTGTGGCAGTGCTCCGCTGCACAGT 3485
      ::::::::::::::::::::
1049 .....SerSerSerThrLeuValAspTyrThrSerThrSerSerThrG 1063
      ::::::::::::::::::::
3486 GCCTGAGCACTGCTGCTTGTACTAATGCTTGAATTAAGCCCAAGTTTG 3535
      ::::::::::::::::::::
1063 LyGIserProValArgIysSerGluGluIysThrAspThrIysArgThr 1079
      ::::::::::::::::::::
3536 GTGGTCGCAATATTAAGAGAGCGTCTGCAG... 3567
      ::::::::::::::::::::
1080 ValIleIysThrMetGluGluTyrAsnAsnAspAsnThrAlaProAlaG 1096
      ::::::::::::::::::::
3568 .....ATGAGAAATGTCAGAAATCAATGATGATGCTTCCAA 3605
      ::::::::::::::::::::
1096 uAspValIleIleMetIleGluValProGlnSerIysTrpAspIysAspA 1113
      ::::::::::::::::::::
3606 AGCCTACCTGCAGAACCAAGCTTAAGCTGTGAAAAAGAAAGAAAAAGT 3655
      ::::::::::::::::::::
1113 sPheGluSerGluGluGluAspValIysThrThrGlnProIleGlnSer 1129
      ::::::::::::::::::::
3656 CTAAACACAGTGAATAAGAAAGAC... 3678
      ::::::::::::::::::::
1130 ValGIysProSerSerIleIleIysAsnValThrThrIysProSerAl 1146
      ::::::::::::::::::::
3679 ...ACCAAGAGACGAGTGTGTGAAGAACGTGTG... 3711
      ::::::::::::::::::::
1146 aThrAlaIysTyrThrGluIysGluSerGluGlnProGluIysLeuGlnL 1163
      ::::::::::::::::::::
3712 .....GACTTACGTACGAAACCT... 3729
      ::::::::::::::::::::
1163 yLeuProIysGluAlaSerHisGluLeuMetGlnHisGluLeuArgSer 1179
      ::::::::::::::::::::
3729 ..... 3729
      ::::::::::::::::::::
1180 SerIysGlySerAlaSerSerGluIysGlyArgAlaIysAspArgGluH 1196
      ::::::::::::::::::::
3729 ..... 3729
      ::::::::::::::::::::
1196 sSerGlySerGluIysAspAsnProAspIysArgIysSerGlyAlaGlnP 1213
      ::::::::::::::::::::
3730 .ACCCATTCAGCAAGAGAGATCCGCCCAAAAGAACACTAGTGTGAGC 3778
      ::::::::::::::::::::
1213 roAspIysGluSerThrValAspArgLeuSerGluGlnGlyHisPheIys 1229
      ::::::::::::::::::::
3779 CTCCTCCACGAAAGCCGTCGAGGAAAGAGTGAAGAGGATGTCTCG 3828
      ::::::::::::::::::::
1230 ThrLeuSerGlnSerSerIysGluThrArgThrSerGluIysHisGluSe 1246
      ::::::::::::::::::::
3829 GCCCTGTGGCCTGAATCCAAACAGCACCACTCAGCTTCAGAGAGTC 3878
      ::::::::::::::::::::
1246 rValArgIysSerAsnIysAspPhe... 1256
      ::::::::::::::::::::
3879 AAGCAAGAGAGTCTCCAGCCAGCAGCTGTCATCCCGCTCAGCCACCTA 3928
      ::::::::::::::::::::
1286 hrProGluArgAspIysIysValAspTyrAspSerArgAspTyrSerSer 1272
      ::::::::::::::::::::
3929 CTACAGGACCCCAAGAAAA... 3948
      ::::::::::::::::::::
1273 SerIysArgArgAspGluArgGlyGluLeuAlaArgArgIysAspSerPr 1289
      ::::::::::::::::::::
3949 .....GAAGTTCCCAAAACCACTCCAGTGA 3974
      ::::::::::::::::::::
1289 oProArgGlyIysGluSerLeuSerGlyGlnIysSerIysLeuArgGluG 1306

```

```

3975 GCCCAAGAAAAAGCAGCCT... 3993
      ::::::::::::::::::::
1306 LuArgAspLeuProIysIysGlyAlaGluSerIysIysSerAsnSerSer 1322
      ::::::::::::::::::::
3994 ....CCACCACAGAAATCAGTCCAGACGACGAAACAGAAAAAGTG 4038
      ::::::::::::::::::::
1323 ProProArgAspIysIysProHisAspHisIysAlaProTyrGluThrIy 1339
      ::::::::::::::::::::
4039 GCTCCCGCCGCCAAGTATCCCTTAACAAAA...CCAAAAAGAAAAAGA 4085
      ::::::::::::::::::::
1339 sArgProCysGluGluTyrThrIysProValAspIys... 1350
      ::::::::::::::::::::
4086 AAAACCA...CCGCCGTCATTAAGCAGAGAGAAATCGACGCA 4123
      ::::::::::::::::::::
1351 .....AsnSerGlyIysGluArg 1356
      ::::::::::::::::::::
4124 CTTTAACATCTCAGCAGCTCTCCAAATGCAATAGTTCTTAAGCAAAA 4173
      ::::::::::::::::::::
1357 .....GluIysHisAlaAl 1361
      ::::::::::::::::::::
4174 ATTCAGCAGATGAGTCCACAGATCAGAGTGAAGTGAAGAGAGATTG 4223
      ::::::::::::::::::::
1361 aGluAlaArgAsnGlyIysGluSerSerGlyAlaAsn... 1373
      ::::::::::::::::::::
4224 TGAACGAGAAATGTGTGGAGATGGAGGCTTAAGAAATCTTGACTCTG 4273
      ::::::::::::::::::::
1373 ..... 1373
      ::::::::::::::::::::
4274 TTCCTATTAACCCAGCGGTGCTTCTCTCTGTGCCAGTAGTGCGCAT 4323
      ::::::::::::::::::::
1374 .....CysHisVal...Tyr 1377
      ::::::::::::::::::::
4324 GTAGAGTTTGTGTATTTGCCAAGCTGTGTGTGAGCCCTTCCACAAATTTG 4373
      ::::::::::::::::::::
1377 rLeuThrArgGlnThrLeuPro.TTPArgArgSerTyrIleuGlnIy 1392
      ::::::::::::::::::::
4374 TTTAAGAGAGAACGAGCCGCTCTGTGAGACCACTGGAATAATTTGGT 4420
      ::::::::::::::::::::
seq_name: /cgn2_6/protodata/1/lna/5A_COMB.seq.us-08-764-233A-1
seq_documentation_block:
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neft, Snezana
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: 1506/CIP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 49377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Sorangium cellulosum
IMMEDIATE SOURCE:
CLONE: p98/1, pJL3, and pVKM15
FEATURE:
NAME/KEY: misc_feature
LOCATION: 927..19874
OTHER INFORMATION: /product= "Sora"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs that
OTHER INFORMATION: are known to be involved in the synthesis of polyketide
OTHER INFORMATION: Saccharopolyspora erythraea."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 927..19874
OTHER INFORMATION: /product= "Sora"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs that
OTHER INFORMATION: are known to be involved in the synthesis of polyketide
OTHER INFORMATION: compounds."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 942..7115
OTHER INFORMATION: /product= "Module 1 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7203..12884
OTHER INFORMATION: /product= "Module 2 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13455..19616
OTHER INFORMATION: /product= "Module 3 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19871..46318
OTHER INFORMATION: /product= "Sorb"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs genes
OTHER INFORMATION: to type I PKs genes"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19870..24556
OTHER INFORMATION: /product= "Module 1 of Sorb"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 24638..30820
OTHER INFORMATION: /product= "Module 2 of Sorb"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30881..35446
OTHER INFORMATION: /product= "Module 3 of Sorb"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 35528..40114
OTHER INFORMATION: /product= "Module 4 of Sorb"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 40190..46318
OTHER INFORMATION: /product= "Module 5 of Sorb"
FEATURE:
NAME/KEY: misc_feature

```

; LOCATION: 46851..47891
; OTHER INFORMATION: /product= "Sorm"
; ; OTHER INFORMATION: /note= "The protein encoded by the sorm gene is highly
; ; OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
; ; OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
; ; OTHER INFORMATION: polyketide rapamycin."
US-08-764-233A-1

alignment_scores:
    Quality: 337.00      Length: 1575
    Ratio: 0.499         Gaps: 82
    Percent Similarity: 42.921   Percent Identity: 21.841

alignment_block:
US-09-B11-045A-1 x US-08-764-233A-1 ..

Align seg 1/1 to: US-08-764-233A-1 from: 1 to: 49377

26 ASPAGLGLATYRALALIEGLYLys.....GluLysProPr 39
||||| ||| ||||| ||| ::| |||||
7113 GACTCGCTGCCCGAGCCTTCGGACAGAGGTCTCCCGCAGCAGCGAGG 7122
39 ophLeuProGluGluProSerSerSerSerGluAspAspProIle. 55
||||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
7163 CGCGCGCTCCGCGCTCTGGCGCCACCAGCACGAGCAGACGCAGATTG 7212
56 ProAlaGluLeuLeuCysLeuIleCysLysAspIleMetHrrAspAlaVa 72
7213 CCATCGTGGCATAGCCCCCTCGTCCGCGGCGGCAATTGGCGATGTGCAC 7262
72 ValIleProCysCysGlyAsnSerSerCysAspGluCysIleArg... 87
7263 GCCTTT.....GGATTCTCTCCACCAAGACGCGAGCGGTGCA 7303
88 .....ThrThrLeuLeu 91
7304 GCCATCCCACTACCCGATGGATGCCGGTGCCCTCAGACCCCTGATC 7353
92 GluSerAspLysHisThrCysProThr.....CysHisGlnAs 104
7354 CCGACGCCAAGGCCAAGAGCTACGTCGGCATGCTGCCATGCTCGACACG 7403
104 nasPValSerPro.....AspAlaLeuIleAlaAsnLysPheLeuArg 119
7404 GTGATCTCTTCGACCTCGATTTCTTG..GCATCACCCTCGTAGAGC 7450
119 InAlaValAsnAsnPheLysAsnGluThrGlyTyrrLysArgLeuArg 135
7451 CA.....AACACATCGACCCCACGACCGCTCACTCCTCGAACTGCT 7494
136 LysGln.....LeuProProPheLeuPheLeuVa 145
7495 GGCAGCGCCTCGAAGAGCGGATGCTCCCTCACCCTCAAGAGATTCT 7544
145 lProPro..... 147
7545 CCCACCGCGCTGTCGTCGGCATGGCGCCAGCAGATAGCGCCGCGGA 7594
148 .....ProArg 149
7595 ACCGGCGCGAGGATTCGNAAGCTTACATGTCCAAAGCACTTAACGCGT 7644
150 ProLeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuAr 166
7645 CTTTTCGGCGGGCGCTTGCGCTTCACGCTGACGTCGCAAGGCGCACG 7694
166 gGlnGlnAspProValAlaPheArgTyrrLysSerProThrCysSerA 183
7695 CTCTCGGTCAACACCGCTTGCTCTCTCGCTGCTGCGCTCCACCTGC 7744
183 spThrLys.....ThrAlaGlySerCysSerAspSeryl 194
:::||| ||| ||| ::|||

```

```

7745 CTGCCAGGCGCTTCGCCGCGATGATGCAACCTCGCCTCGCCGCGAGGG 7794
195 ThrLeuSer...ArgLeuProAlaProSerIleSerSerLeuThrSerAs 210
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7795 CCGTCGTATGTCCTCCGAGACCTTCG...TCCCTCTTCCCGCC... 7839
210 nGlnSerSerLeuAlaProProValSerGlyAsnProSerSerAlaProA 227
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7840 ...TGCCGCGCTTGGCCCGCCGACGCGCTCCAGACCTTCCTCGCGCAG 7885
227 IaproValProAspIleThrAlaThrValSerIleSerValHisSerGlu 243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7886 CGCCG...ACGGCTACGGTGGGCTGAGGCTGACGCTCATGCTC 7923
244 LysSerAspGlyProPheArgAspSerAspAsnLysLeuProAlaAl 260
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7924 ...TTGCCCTCGAG 7934
260 aAlaLeuThrSerGlnHisSerLysGlyAlaSerSerIleAlaIlePhe 277
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7935 CGGCTCCGCGACGCGCTTGGCCCAAGAGACGCGCTCTCGCCGCTGTCG 7984
277 IalaMetGluGluLysGlyValProGlyThrSer... 288
7985 CG...GACCGCGCTCACACCGACGAGCGCGCATCGACG 8019
289 ...Pro.TyrAsnSerIlePheValGlyLysLeuLeu 301
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8020 GCATCCAGCGCCGCCAATGGCAC... 8041
301 IsglyGlnLeuIleProThrThrGlyProValArgIleAsnAlaAlaArg 317
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8042 ...CTCCAGAGAGAGGCTCTCCGCGCGCGCTCCACAGACG 8080
318 ProGlyGly...ArgProGlyTyrGlnHis 328
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8081 CCGCATCGCTCCCGCGATGTCGACGCTGTCGATGCCATGG...CACCG 8127
328 rAsnLysLeuGlyTyrLeuValSerPro... 338
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8128 GCACCTCTTGGGGATCCCATCGAGGTCCAAAGCCCTGGCTGCTGTAC 8177
339 ...GlnGlnIleArgArgGlyLysArgSer...Cys... 348
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8178 GGTGAAGGCACATC...CGCGAAAAGCCTCTTTCTGCGTGCCTGCA 8224
349 ...TyrArgSerIleAsnArgGlyArgHisHis...SerGluArgSerGln 363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8225 GACCAACGTGGCCACCTCGAAGCGCGCGCTCGGGGCTCGGCGCA 8274
364 ArgThrGlnSerProSerLeuProAlaThrProCysPheValProValPr 380
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8275 AGATCGTGGCTCCCTCC...TCCACACGCGCTCGCC 8309
380 oProProProLeuTyrProProProHisThrLeuProLeuProPro 397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8310 CCCACCC...TCCACACCAACCCACGCAATCCCTGATGCGCTG 8350
397 LysAlaProProProGlnPheSerProGlnPheProSerSerGlnProPro 413
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8351 GGATGCGCTCCCGTCG...CCGTCGTGATGGCCAGCA 8385
414 ThrAlaGlyTyrSerValProProProGlyPheProAlaProAlaAs 430
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8386 GGCCTTGGG...TCCGCCACGCGGATGGGCTCCCGCGCGCGCGCG.. 8430
430 nIleSerThrAlaCysPheSerProGlyValProThrAlaHisSerAsn 447
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8431 ...TCTCGGCTTCGACTCTCG...GCACCAAGCTCAAGTCATCTCT 8473
447 hrMetProThrThrGlnAlaProLeuLeuSerArgGluGluPheTyrArg 463
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8474 CGAAGAGGCGCCCGCATCGCCGAGGTGAGCGCGCATCGACAGCGCGG 8523

```

```

464 GluGlnAsnAspLysGlyArgGluSerLysPheProTyrSerGly... 478
8524 ...CGTCCGAGCGCGCTCCCGAGCGCGGCGCGCG 8555
479 .SerSerTyrSerArgSerSerTyrThrAspSerSerGlnGlyLeuAlaG 495
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8556 CTCCGTGCGGCCAAGAGCGAGCGCGCGTGGCGCGCGAG...C 8596
495 IohisIleHisAlaLeuThrLeuSerProSerAlaAlaHisThrLeuAsp 511
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8597 AAAGCGGCTCCGCGACCACTCTCTGCGCAAAAGCAAGCTCGCCGCGCG 8646
512 Leu...LeuHisAspHisProHisProProGluAlaGluAlaAr 526
8647 ACGTGCCTATTCGCTCGCGACACGCGCGCCACTTCGAGAGCGCG.. 8694
526 gSerIleMetIleValHisMetProAspLeuMetAspIleAlaHisAla 543
8695 ...CCGCTCTCTCTCAAGCAAGCGCGCA 8719
543 rSerArgSerProProTyrArgArgTyr... 552
8720 CGAGCTCTCTCCGCTCGATTCGCTGGCCCAAGACATTCGCGCGCG 8769
553 ...ArgSerArg.S 556
8770 TGCTCGACGAAGCGCGCCCGCAAGAAAGCTCGCCTCTCTCAAGGG 8819
556 eArgSerProProGluPheArgGlyLysSerProThrLysArg... 570
8820 CAAGGAACCCAGCGCGCCACCATGGGCGCGCGCTCTCGACGCTTTC 8869
571 AsnValProArgGluGluLysGluArgGluTyrPheAsnArgTyrArgG 587
8870 CGCTTCGCGGAGCGCTTCGACACCGTGGCGCGCCACCTCGACCGCG 8919
587 uValProProProTyrAspIleLysAlaTyrTyrGlyArgSerValAsp 604
8920 TCGACCGCGCCCT...G 8933
604 heArgAspProPheGluLysGluArgTyrArgGluTyrGluArgLysTyr 620
8934 CGCGACGCTCTCTCTCGCTCCGACGCGCTCGCA... 8965
621 ArgGluTyrTyrGluLysTyrTyrLysGlyTyrAlaValGlyAlaGlnPr 637
8966 ...GCAAGCGCGCGCGC 8979
637 oArgProSer... 640
8980 TCGACCAAAACCGCTTCACCGACGCGCGCTGTGGCCCTCGAAGTCGCC 9029
641 ...AlaAsnArgGluAspPheSerProGluArg 650
9030 CTCTTTCAGCTTCTCCAAATCCCTCTGAAAGCCCGCTCTCTCTCTG 9079
651 LeuLeuProLeuAsnIleArgAsnSerProPheThrArgGlyArgArgG 667
9080 ACACCTCAATGGCGAGCTCGCGCGCA...CGTGGCGGCGG 9120
667 uAspTyrAlaAlaGlyLysSerHisArgAsnArgAsnLeuGlyAsnT 684
9121 TCTTTCCTCCAGAGCGCTCGACCTCTCGCGCG... 9156
684 yrProGluLysLeuSerThrArgAspSerHisAsnAlaLysAspAsnPro 700
9157 ...CCGCGCAAGCTCATGCAAGCGCTCCACAAAGCGCGCG...CCA 9198
701 LysSerLysGluLysGluSerGluAsnValProGlyAspGlyLysGlyAs 717
9199 TGGTTCACCC... 9207

```

```

717 nlySHSLySLySHSLySArGLySArGLySArGLySGLyGLuGLuS 734
7208 .....TCGAGACCTCGAGAGAGAGT 9229
734 erGLuSerPheLeuAsnProGLuLeuLeuGLuThrSerArGLyScyArG 750
7230 CCGGACACCTTCACAGCCCTACAGATGAGACAGTACGCTCGCCG 9273
751 GlySerSerGLyLeuSpgLUThrLysThrAspThrLeuPheValLeuPr 767
9274 CCTCAATGGGCTCTCTCCACCGCTGTCG.....CTGGCATGAA 9314
767 oSerArgAspAspAlaThrProValArgAspGLuPrometAspAlaGLuS 784
9315 GACGGGTGTGAGATCGCCCGCAGGCCGAGACCTCGGAGAAAGAC 9364
784 erLThrPheLysSerValSerAspLysAspLysArGLySAspLys 800
9365 CACACGCTGGCGGTCCAGCACGCTTCACCTCCGACATGAGAGGAA 9414
801 ProluValLysSerAspLysThrLysArGLySArGLySerAspLysAlaThr 817
9415 TCGTCAGACACTTCGCCGGGTCCAGGCTCACTACATCCCGCA 9464
817 rAlaLysLysAspAsnValLeuLysProSerLysGLyProGLuGLuLysV 834
9465 CGCATCCCATCATCTCCAAAGTACCGGGCGCGCCACAGG..... 9507
834 aLAspLysAspArgLysSerProArg..... 843
9508 .....ACCAGAGCTGCGCTCGCCGACCTACTGGGTCCGCCACG 9546
843 ..... 843
9547 TTCGCCACACCGTCCGTTCTCGACGGCTACGTGCCCTTCAGCGCGAA 9596
844 .....SerGLuProL 848
9597 GGGGACAGGCTTTCTCGAGCTGGGCTTCACGCTGTCTCTCCGCCCT 9646
848 euLysLysAlaLysGLuGLuAlaThrLysLysSerValLysProSer 864
9647 TGGCAAGAGCGCTCGACAGAGCAAGGACAGTGGCATGGCCCTCC 9696
865 ...SerSerSerGLuLysAspLysValThrGLyThrProArgLysAl 880
9697 TTCACACCTTCGCAAGGAC.....GCGACGACCGCCGAGGGCTTC 9737
880 aHISerLysSerAlaLysAspThrArgArgLysSerGLuPro..... 894
9738 ACCGCGCGCTGGCGCTCTCCACGCTGACGGCTCACACCCGACTGGAG 9787
895 .....ArgThrArgSerLysArgThrVal 903
9788 CGCTTCTTCGCCCCCTCGCTCCAGCAAGGTCTCCCTCCACACTAG 9837
904 ProluThrSerSerGLuLys...SerGLuProValArgThrArgArgPr 919
9838 CTTTCAGGCGAGCGCTTCTGGCTGATGCTTCGCAAGGACACGCTCC 9887
919 oArSer.....LeuArgLysLysAsnTyLLeuLAla 931
9888 GACGTGGCTCCGCGAGCTGACCTGACCGATACCCGCTGCTGGCGC 9937
931 rGLuLysAsnGLuArgGLuLysArGLySArGLySerValAspPhe 947
9938 CGCGTTCCTCCGCGACCGCATGCTCTCTCCACAGACGACTCT 9987
948 GLuSerSerSerMetLysLysLysValGLuGLyThrGLuLysValLys 964
9988 CACTCTCAGACATCCGTGGCTCGCCGATCAGCTGCTTCG...GTACA 10034
964 sProSerProluArgLysMetGLuLysArValGLuLysLeuGLuArgT 981
10035 CCACTCTTCGCG.....GCACCTGCTTCTCGAGCT 10066
981 hrProGLuLysAspLysLysLysSerSerThrThrProAlaLysLysL 997
10067 TGGCTGCTGTCGCGCGGTGCGCTCGGCTCGACACCGTGGAG..... 10110
998 LysLeuAsnArgLUThrGLyLysLysLysLysAlaLysAlaLysAlaL 1014
10111 .....AGCTC 10115
1014 rThrThrLysGLuProSer...GLuLysLeuGLuSerThrSerSerLysT 1030
10116 ACCCTGAAACCCCTCGCTCTCCGCTCGATGAAGGCCCTCTCGCC 10164
1030 LeuLysGLuGLuLysValLysGLyLysAlaLysArGLySValaLys 1046
10165 ...AGGTGTGGGTGGGCTTTGGACAGCAGGACGAG..... 10201
1046 rGLuGLySerSerSerThrLeuValAspLysThrSerThrSerSerTh 1063
10202 .....GCCACTCTCTTCAC 10217
1063 LysLysSerProValArgLysSerGLuGLuLysThrAspThrLysArgTh 1079
10218 ACCCGACCCAGAGCGCTCTCTCAGAGCGCCCTTGACTGCCACGGAG 10267
1079 rValLysLysThrMetGLuLysThrAsnAspAsnThrAlaProAlaG 1096
10268 C.....GGCTGCTGCTCCAGCT. 10286
1096 LuAspValLysLysLysLysLysLysLysLysLysLysLysLysLys 1112
10287 .....ACCCGCTCCCTTCCTTCGATCTTCAC 10313
1113 AspPheGLu.....SerGLuGLuGLuAspValLys...ThrThrG 1125
10314 GACTGGCTCTCTCGGGCGCCACCGATGACACCCAGGCTTCACG 10363
1125 LuProLysLysValLysLysProSerSerLysLysLysLysLysLysL 1141
10364 AACCTCGAAGAGCGCTGCTGCTTCAGCGCTTCAGTTCAGGGCTCC 10413
1142 ThrLysPro.....SerAlaThrAlaLysThrGLuLysGLuSerGL 1156
10414 GCTGGTGTGAGAGCGCGGACGAGCTTCGGAAGCTCAGCTCCG 10463
1156 uGLuProGLuLys..... 1160
10464 GACGCCGCCAAAGATGCCGCTTGCCCTGACCCGCCGCTGCT 10513
1161 .....LeuGLuLysLeuProLysGLuLysLysLysLysLysLysL 1172
10514 CGACAGGCGCTCGACGCGCTTCCTTCGACGACGAGGGGACAGCGG 10563
1173 MetGLuHISLysLysLysSerLysGLy..... 1182
10564 TCGCGTGGCTTCCTGCTGGGGGAGTCTCTGCGCGCTGCTGCTCC 10613
1183 .....SerLysSerGLuLysGLuArgAlaLysAsp 1194
10614 ACACAGCTGGGCTGCGCTTCACACCGTCCGAAAGCGGAAAGCGGCTC 10663
1194 rGLuHISerGLySerGLuLysAspAsnProAspLysArg..... 1207
10664 GCTGTCTCTGCGCGAGCGCGAGGAGCCATGCTCGTCCGTCAGAGCG 10713
1208 ...LysSerGLuLysLysLysLysLysLysLysLysLysLysLysL 1223
10714 TCGCCAGGCGATACGCTCGCGAGAGCTTCGCAAGGAGGCTTC 10763
1223 rGLuGLuGLuHISerPheLysSerSerLysSerLysGLuThrArg 1239

```

```

10764 C.....ACCAGATGCGCTCTTCGCGCTGACTGAGCGCA 10788
1240 ThrSerGluLysHisGluSerValArgLysSerAsnLysAspPheTh 1256
      :::::||||| :::::||||| :::::|||||
10769 GGTGGCGAGCCCGACCTCAGCTCTGAGAGCCCGACGCGCT..... 10840
1256 rProGluArgAspLysLysValAspTyrAspSerArgAspTyrSerS 1273
      |||:::||||| :::::|||||
10841 .CCTTCTGGGCAATCGAGCGGCTCGACCTCGCCCGAGGTGCTCTCGCC 10889
1273 eLysArgArg.....AspGluArgGlyLuuLeuAlaArgArg 1285
      :::::||||| :::::||||| :::::|||||
10890 GCGGCTGGCGACCTCGCTGCCCTCGACAGCGCGCTGA...CCAAGTCC 10936
1286 LysAspSerProArgGlyLysGluSerLeuSerGlyLuuLysSerLy 1302
      ||| :::::||||| :::::|||||
10937 TTGCGCTCAGGCGCTGCTGCTGCTCCCTCATGGCTAGAACCGCGAG 10986
1302 sLeuArgGluGluArgAspLeuProLysLysGly.....AlaG 1315
      : :::::||||| :::::|||||
10987 ACCATCATCAGAGCGCCCATCTCATCGCGCGCGCGCTCGCTGCTG 11036
1315 LuSerLysLysSerAsnSerPro.....ProArgAspLys 1327
      ||| :::::||||| :::::|||||
11037 CAAGCGTGGCTGGCGAGAGACGCGCTCGCTCGCTGCTGCTGCT 11086
1328 LysProHisAspHisLysAlaProTyrGluThrLysArgProGlyGlu 1344
      ||| :::::||||| :::::|||||
11087 CACCGGAGCGCGCATCGCTGCGCGCGCGATGAGAGCTCAAGGACCTG 11136
1344 uThrLysPro 1347
      :::::|||||
11137 CTCACGCCCC 11146

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-144-085-3

seq_documentation_block:
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziemann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144, 085
; EARLIER FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-144-085-3

alignment_scores:
Quality: 329.00 Length: 1717
Ratio: 0.452 Gaps: 76
Percent Similarity: 42.400 Percent Identity: 19.802

alignment_block:
US-09-811-045A-1 x US-09-144-085-3/rev ..
Align seg 1/1 to reverse of: US-09-144-085-3 from: 1 to: 33529
111 LeuIleAlaAsnLysPheLeuArgGlnAlaValAsnAsnPheLysAsnG 127
||| |||:::||||| :::::|||||
::: :::::|||||

```

```

8324 CTCCTGTGCTACCGGCTCCCTGCGTCAACACCGAGCGG...CGCATCGA 8278
127 uThrGlyTyrThrLysArgLeuArgLysGluLeuProProPheLeuPhe 144
      | :::::||||| :::::|||||
8277 GCCAGCGCTACACCGAGCGGCTCTGCGCACAGCG...CAGCACTCTGATC 8234
144 euValProProArgProLeuSerGlnArgAsnLeuGluProArgSer 160
      ||| :::::||||| :::::|||||
8233 ACTCGCGAGGCGCTTCGCCCTTCCTCCAGCGGACCTGCTCCAGCGCG 8184
161 ArgSerProIleLeuArgGlnAlaAspProValValPheArgTyrThVa 177
      ||| :::::||||| :::::|||||
8183 CTCACACCTCTGACCGCGCACGAGCTCGCC...TCCAGCCCTCTCAACT 8137
177 LysProThrCysSerAspThrLysThrAlaGlySerCysSerAspSerG 194
      :::::||||| :::::|||||
8136 GCGCGCGAGCTGCGCTGCTCTGCGTGGCGCGCTCG... 8100
194 LysThrLeuSerArgLeuProAlaProSerIleSerSerLeuThrSerAsn 210
      :::::||||| :::::|||||
8099 .ATGATCAGCTCTGACCGCGAGCTCCGAGGCGATCACTCACTC... 8058
211 GlnSerSerLeuAlaProProValSerGlyAsnProSerSerAlaProAl 227
      :::::||||| :::::|||||
8057 ...GCCGCGGTACGCGCTTCGCGCACGCGAGATGACACGAGGCAACT 8011
227 aProVal..... 229
8010 GCCCTCAGCTCGCGCTTCCCTTGGACGCGTGGCTTCCCTTGGCACG 7961
230 .....ProAspIleThrAlaThrValSerIle..... 238
7960 COTCTGTGTAACACGACGACGACGCGTGGCTTGGCTTCCCTTGGCG 7911
239 .....SerValHisSerGluLysSe 245
7910 CGCCAGCGCGGAGGTACGAGGACGAGCGCGGAGCTCTCGCGACGCTG 7861
245 rAspGlyProPheArgAspSerAspAsnLysLeuLeuProAlaIleAla 262
      |||:::||||| :::::|||||
7860 GTCAAGC.....AGCTCAGAGATCCGCAACGCTTCCCTTCTCT 7820
262 eThrSerGluHisSer.....LysGlyAlaSerSerIleAlaIleThr 276
      |||:::||||| :::::|||||
7819 GCACCGGCTCCACAGACGACGCTCGCGCACAGAACATGCGCTGCTGCC 7770
277 AlaLeuMetGluGluLysGlyValProGlyThrSerPro..... 289
7769 GCGGTC.....GTCTTGGGCGCTCCACGACGATAGCGCTG 7735
290 ....TyrAsnSerIlePheValGlyGlnSerLeu..... 299
7734 CCGCTGGAACGCG...TACGTGCGAGCTCCACGAGGCGCGCGCACGCG 7688
300 .....LeuHisGlyGlnLeuIleProThrThrGlyProVal 311
7687 CCGACAGCACCTGGGCGCCAGTCACTGCTGCGCTCGACGATGCACTGTC 7638
312 ArgIleAsn..... 314
7637 GCCACCGCGCTCGCAGGCGCGCTCTTCCTGCTGCTCGCGAGGCT 7588
315 .....AlaAlaArgProGlyGlyArgProGly...TyrG 326
7587 CGTCACAAAGCTCGCTGCGCTCTCGGAGGACGACCGCGCGCGCACG 7538
326 LuHisSerAsnLysLeuGly.....TyrLeuValSerPro... 337
7537 CGCACAGACGACGACATCGGCGCCACACTGAGAGTATGTCCTCACCCGCC 7488
338 .....ProGlnGlnIleArgArgGlyGluArgSerCy 348
7487 GCCGCAAGCGGTGCGCATCCGCTCCAGGAGGACGACCGCTTCCGACCTG 7438

```

| | | | | |
|------|--|---------------------------------------|-----------------|------|
| 348 | styr..... | ArgserLleasnArggly | ArgHisHisSerGlu | 360 |
| 7437 | CCGACACCGAGTACTGCGCGACATGAGGGCTTTCTGCGACCGAGGCTGC | | | 7388 |
| 361 | ArgSerGlnArgThrGlnSerProSerLeuProAlaThrProCysPheVala | | | 377 |
| 7387 | CCGTCACGCGCGCTCCACACCGCGACCGGTGGCGGGGTACTGTC..... | | | 7343 |
| 377 | lProValaProProProProPoleuTYrProProProProHis..... | | | 390 |
| 7342 | ...ACCCGCGCGCTCCCTCCCGAATCTTCAGACATCCCGTCATATGTC | | | 7297 |
| 391 | | ThLeuPro | | 393 |
| 7296 | CGCGCTGTGGAAAGCGGTGCACACACGACAGACGCCGGGTGGCGCGCCCT | | | 7247 |
| 394 | LeuProProGlyValaProProProGlnPheSerProGlnPheProSerSe | | | 410 |
| 7246 | GGCGCTTCAGCGCTGCGCGCACCGGACACCGCGCGCTTCGTCGCCGCTC | | | 7197 |
| 410 | r..... | | | 410 |
| 7196 | ACGACCCGTGCATGGCGCGTTAGCCCCGGCATGCTCATGCCCCCTG | | | 7147 |
| 411 | ...GlnProProThrAlaGlyTYrSerValProProProGlyPhePro | | | 425 |
| 7146 | CGCCCCGACCTCGACAGCGCGCGCTGCATACATCCGGGCTCGAGGCTGCA | | | 7097 |
| 426 | ProAlaProAla.Asn..... | | | 430 |
| 7096 | CCGACACCATTCGCTCCCCGGCTCGACCCCTGCATCAGCGACGCGG | | | 7047 |
| 431 | | Ile | | 431 |
| 7046 | GGCGACACTAGCTTGGTGGCTGCGCGACAGGCTGCAGACACCGGCACGTG | | | 6997 |
| 432 | SerThrAlaCysPheSerProGlyValaProThrAlaHisSerAsnIhme | | | 448 |
| 6996 | CGCAGCGCTCATCTCTCATGATGCGCCMAAGMAGCGCACCGGGATTCA | | | 6947 |
| 448 | rProThrThrGlnAlaProLeuLeuSerArgGlnIuPheTYrArgGlu | | | 465 |
| 6946 | GCCC...CMAGACATCCCACTGACGAGTGCACAGCGCACATTCACGCGCAAC | | | 6900 |
| 465 | lAsnAspIysGlyArgGluSerIysPheProTYrSerGlySerTYr | | | 481 |
| 6899 | AG.....CCCGGCTGGATGATCTCGTCCGTCACACAGCGTCCTTC | | | 6856 |
| 482 | SerArgSerSerTYrThAspSerSerGlnGlyLeuAlaGlnHisIleHt | | | 498 |
| 6855 | CTC..... | | | 6824 |
| 498 | salaleuThrLeuSer.....ProSerAlaAlaHisThrLeuAspL | | | 512 |
| 6823 | TCACACCGACGCTGCAGATAGCGGCTGCACGCGCTGCGACACTCGTGCAC | | | 6774 |
| 512 | euleuHisAspHis..... | ProHisPro... | | 519 |
| 6773 | GGCGACGGACACACGGGTACACTCTGTAAAGCGCTCTCCCATCCCGAG | | | 6724 |
| 520 | | ProGluGluAlaGluAlaArgSerAlaMetLleVala | | 531 |
| 6723 | CGCGTGGCTGCCCTGCCGCTAAGACAGACCGCAAGCTTCCGCGCTGCT | | | 6674 |
| 531 | lHisMetProAspLeuMetAspLleAlaAlaHisAlaArgSerArgSerPro | | | 548 |
| 6673 | TCGCGCTGCC..... | GCATACACACGCGCGCATTCGCGCGCG | | 6639 |
| 548 | roTYrArgArg..... | TYrArgSerArgSerArgSerPro | | 559 |
| 6638 | CCCGACAGAGGCGCGGAAGACCTTCACAGTCTCCGACGAGGCTGCGCGC | | | 6589 |

[illegible]

```

5775  |||  ::::  |||
5775  GCCTGCCAGAGAGCTTTCAGACGCGCTCTCGGCGCATCCCTTGAG  5726
786  .....  ThrpheylserValSerAspLysArg  796
5725  CCGGCTGAACTCCAGAAACGGGGGTCATCCGTCACCGCCGC  5676
797  GILuysApLysProLysValLysSerAspLysThrLysArgLysSerAs  813
5675  CGGCCAGCGGCGAGTCCGATTCACCCGTACGCAACGCCGTGCAGCGCAGG  5626
813  pGlySerAlaThrAlaLysLysAspAsnValLeuLysProSerLys...  828
5625  TGCAGCGACACAGCGACGACGACGCGCTGCCACCGTCATCGCTGG  5576
829  .....  GlyProGlnGluLysValAspLysAspArgGluLysSerPro  842
5575  GCCATGCGACGCCGACGACGATAGGCCACACGCGCTGAGAGCAGCGCTCCG  5526
843  ArgSerGluProProLeuLysLysAlaLysGluGluAlaThrLysIleAs  859
5525  CGCTCCCGATGCCCGCGGTACCGCTCCGACATCTCCAGCCCGC  5484
859  pSerValLysProSerSerSerSerGlnLysAspGluLysValThrGlyT  876
5483  .....  CACCTGCTGAAGCGCCATAGTCGAAACCGCGGCCCCAGATA  5441
876  hrProAdgLyAlaHisSerLysSerAlaLysAspThrArgGlnSer  892
5440  CACCCCGG...  TGGCGCTCCGCTCAGCGCGGCGGCGCACGCCGCGCAC  5394
893  GlnProArg.....  895
5393  GCTCCATGCTCCCATCGCGCTCCAGACACGCCGCTGGGAGATCC  5344
896  .....  ThrArgArgSerLysArgThrValP  904
5343  ATGCGCTGCGCTCTCGGGCGGATATCCGAAAGTCCGCGTCGAAACAG  5294
904  rOlys.....  ThrserserGlnLysSer  911
5293  GTCGATGTCGCCGAGAAATCCACCCCTCGCGGCTGAGCTCTTGCCCGCAG  5244
912  GlnProValArgThrArgArgProArgSerLeuArgLysIleAsnTyrLe  928
5243  CMTCCGGGTGGGGGTCTAGACCGAAA.....  5217
928  uIleAlaArgGluLysAsnGluArgGluLysArgLysLysSerValAspL  945
5216  ....  GCCTCTCCAGCGCGAGGAGGCCCTCGATCACGTCCTTCCCTC  5171
945  ysaAspPheGlnUserSerSerMetLysIleSerLys.....  ValGlu  958
5170  GGGCAAGAGAGCGCATAGTCTTCCGGCGTCTGACGCGCTCCGCGCAGCC  5121
959  GlyThrGluLeuValLysProSerProLysArgLysMetGluLysAspVa  975
5120  GGCAGCGCATGATAGATAGCAGCGCGCTGCTTCCCGCATGACGA  5072
975  lGluLysLeuGluArgThr.....  Prog  983
5071  .....  CGACCCCGCTCGGGGCTCCCATCAGCTGCGGTCCCG  5034
983  lulysAspLysIleAlaSerSerThrThrProAlaLysIle.....  997
5033  AAAAGGCCAGCTTACGAGACACTCCGATGCGCCGCGGTGGGGTAG  4994
998  .....  Ly  998
4983  TCGAAGACAGCGCTGCGCGGACAGACGTCTCCGCGGCGGAGAGCG  4934
998  sleuAsnArgGluThrGluLysLysIleGlyAsnAlaGluAsnAlaSerT  1015
|||:::  |||:::  :::  |||
4933  GTTCCGAGCTCCACCGCATCAGCGAGTCGAGCCGAGGCCCTTGACA  4884
1015  hrThrLysGluProSerGluLysLeu.....  1023
4883  CTGTCTCGGCCCGCCACGGCTCGCTGCGCTGAGCCCGAGACCGCGCGG  4834
1024  .....  GluSerThrSerLysIleLysGlnGluLysValLysGlu  1037
4833  ACCTCGCCCCCGACACAGCTCATGATAGCGCATTCAGCGCGCTCCGCTCCG  4784
1037  yLysAlaLysArgLysValAlaGlySerGluGlySerSerSerThrLeu  1054
4783  CACGCGCGAGAGCGCGCGCGGACGCGGCGGTCTCGCTCGCTCGCGG  4734
1054  alaSP.....  TyrThrSerThrSerThrGly  1063
4733  TGCAGCGCTTCCGACAGCTGGGGCGCAACAGCCGACGAAACAGCGCGG  4684
1064  GlySerProValArgLysSerGluGluLysThrAspThrLysArgThrVa  1080
4683  AGCTCGCGCTTGGGCTCCAGCCACGCTGCAGCTGCGGAGATCCAGGTG  4634
1080  lIleLysThrMetGluGluTyrAsnAsnAspAsnThrAlaProAlaGlu  1097
4633  CATAGGACACAGCGCTCTCAGGCGGTGACAGCGCGCTCCAGCAGCG  4585
1097  spValIleIleMetIleGlnValProGlnSerLysTyrAspLysAspAsp  1113
4584  A.....  MAgCCCTC  4574
1114  PheGluSerGluGluGluAspValLysThrThrGlnProIleGlnSerVa  1130
4573  TTCACGAGACATGACACAGACTCGGTAGCAGTGAACCGTGCAGCTCCG  4524
1130  lGlyLysProSerSerIleIleLysAsnValThrThr.....  1142
4523  CGGC.....  CCCAGCTGTGCGCCATGCGCACCGACCGCGCGGTCTCAG  4480
1143  .....  LysProSerAlaThrAlaLysTyr  1150
4479  ACACCCCAGGCGAGGCTGTCGCGCGGAGCGCGGCGCGGTCCG  4430
1151  ThrGluLysGluSerGluGlnProGluLys.....  LeuGlnLys  1163
4429  CGCGAACGCGCTCGAGAACGTAATCCGCCGCGGTAAATTCGCGGCGG  4380
1163  sleuProLysGluAlaSerHisGluLeuMetGlnHisGluLeuArgSerS  1180
4379  CATGCCCAACGTAACCGCGCGCGCAGAGAAACAGCAAGCGCGCGCA.G  4331
1180  erLysGlySerAlaSerSerGluLysGluArgAla.....  1191
4330  ATCCAGCTCCCGGTGAGTCTGTCAGAGTGACAGCGCGCTCCACCTTGC  4281
1192  .....  LysAspArgGluHisSerGluLysSerGluLysAspAs  1203
4280  GCGCGACACCCAGCGAGCGCGCTCGCGCGCTGCGTCAATCAACGCCA  4231
1203  nProAspLysArgLysSerGluAlaGlnProAspLysGluSerThrVala  1220
4230  TCATGAGACACGCCAGCGAGTGCAGCACCG...  CGCTCAGCGGCGCGC  4184
1220  spArgLeuSerGluGlnGluHisPheLysThrLeuSerGlnSerSerLys  1236
4183  CGGCTGATGCGGCGCAGACGCGCGGATCTCTCCGCTTCGACACAGT  4134
1237  GluThr.....  ArgThrSerGluLysHisGlu  1245
4133  CGCAGCGCACATCGACACAGGTCTCGCGCGGAGCTCGAGCGCATTTGCA  4084
1245  userValArgLysSerSerAsnLysAspPheThrProGlyArgAspLysL  1262
4083  CAACTCGCGGCGCGCGCGCTCCAGCGCGCGCGCGCGAGCTGACGACA  4034

```



```

1262 yValaSPlyrAspSerArgAspTyrSerSerSerLysArgArgAspGlu 1278
      |||||  ::::  |||  :::::  ::  ::|||
4033 AGTGGCTTACACCCCGGCGCGGACAGATGCGGTACCGCCCTGCCC 3984
1279 Arg..... 1279
      ::
3983 CAGACCGCCACGCGCTCCGGTACCAGCAGCGGTGCCGAGGTCACGCC 3934
1280 .....GlyGluLeuAlaArgArgLys.....AspSerProPro 1291
3933 CCGCGGCTCGGTGAGCTCTCCGCGCGCTGTACGCGCACACCGCC 3884
1291 rGgLYLysGluSerLeuSerGlyGln...LysSerLysLeuArgGluGlu 1306
      |||||  ::::  |||  :::::  ::  ::|||
3883 GGGCGCAGACGCGCGCGCGCGCGCACAGCTCCGCGCTCCGCCCT 3834
1307 ArgAspLeuProLysLysGlyAlaGluSerLysSerAsnSerSerPr 1323
      |||||  ::::  |||  :::::  ::  ::|||
3833 CCGCAGCGCGCGCGCGCAC.....GCCACGCG 3806
1323 oProArgAspLysLysProHisAspHisLysAlaProTyrGluThrLys 1339
      |||||  ::::  |||  :::::  ::  ::|||
3805 TCACGCGGCTCGGTCCCGCACATCGATCAAGCGCGCGCTCGGGGTG 3756
1340 .....ArgProCysGluGluThrLysProValAspLysAsnSerGly 1353
3755 CTCGCTGCGCGCGCGTCCGAATAAGCCCCACAGCGCGCTGCGCAGGT 3706
1354 LysGluArgGluLys.....HisAlaAl 1361
      |||||  ::::  |||  :::::  ::  ::|||
3705 CCGCAGCGCGTGTCCGAGCTGCGCTGACCGCGCGCGCTGACCCAG 3656
1361 aGluAlaArgAsnGlyLysGluSer.....SerGlyA 1372
      |||||  ::::  |||  :::::  ::  ::|||
3655 ACCAGCTCCAGACGCTGAGCGCGCTCCGACACACCGCTTGCAGAG 3606
1372 laAsnCysHisValTyrLeuThrArgGlnThrLeuProThrArgArgSer 1388
      |||||  ::::  |||  :::::  ::  ::|||
3605 CCACAGCGCTGCTGCTGCTGCTGCGACAGACCGCGCTGCGGAGC 3556
1389 TrpLeuLeuGlyArgTrpArgArgAlaProSerSerArgAsnPro 1403
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
3555 GCGGTG.....AGCTGCGCAGCGT.....TCGTGCGATCAACA 3523
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:us-08-258-261B-6
seq_documentation_block:
: Sequence 6, Application US/08258261B
: Patent No. 5639949
: GENERAL INFORMATION:
: APPLICANT: Schnupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/258,261B
: FILING DATE: 08-JUN-1994
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-08-258-261B-6

alignment_scores:
      Quality: 323.50      Length: 1578
      Ratio: 0.495      Gaps: 83
      Percent Similarity: 41.445      Percent Identity: 22.687

alignment_block:
US-09-811-045A-1 x US-08-258-261B-6 ..
Align seg 1/1 to: US-08-258-261B-6 from: 1 to: 28958

118 ArgGlnAlaValAsnAsnPhenylAsnGluThrGlyTyrThrLysArgLe 134
      |||||  ::::  |||  :::::  ::  ::|||
7748 CGGACGCCACCTCCCAACAGCGCGTCTGCGGATGAGCCCGCTGCTT 7797
134 uArgLysGlnLeuProProPheLeuPheLeuVal.....ProProp 148
      |||||  ::::  |||  :::::  ::  ::|||
7798 GCCCTGCGCGCGCTGCGCTGCTGAGACAGACAGACACCGCTGAC 7847
148 roArgProLeuSerGln...ArgAsnLeuGlnProArgSer..... 160
      |||||  ::::  |||  :::::  ::  ::|||
7848 CGTGGCGGACATCGACTGGGCGCGCTTGGCGCTTGCATGAGCGCGCTC 7897
161 .....ArgSerProLeuLeuAr 166
7898 GCCCGCGCGCGCTCTGCGCGATTTGCCGAGGCGGCGCGCTCTCGAG 7947
166 g.....GlnGlnAspProValAlaPheArgTyrThrValSerProT 180
7948 ACCAGGAGAGCGCGCTCTCCGAGCATGGCCGCGCGCGCTCTCTGCA 7997
180 hCysSerAspThrLysThrAlaGlySer.....CysSerAspSer 193
      |||||  ::::  |||  :::::  ::  ::|||
7998 CAAGCTCGGAGCGCGCTGCGAGCAGCAGCAGCTGCTGCTGCTGCG 8047
194 GlyThrLeuSerArgLeuProAlaProSerLysSerLeuThrSerAs 210
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
8048 TGGTGGCCACGAGAGCGGCC.....TCGTC 8073
210 nGlnSerLeuAlaProProValSerGlyAsnProSerSerAlaProA 227
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
8074 CTCGCGCAGAGAGCGCGCTCTCCATGTCG..... 8101
227 laProValProAspLysThrAlaThrValSerLysSerValHisSerGlu 243
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
8102 .....ACCCGACAAAGGCTCTCGATCTCGGTGATTCGATTCATATGG 8146

```



```

9683 CCCTGCTCC.....TGTCGGCCAGGA..... 9703
760 rAspThrLeuPheValLeuProSerArqAspAlaThrProValArqA 777
9704 .....GCGAGCGCGCTCGCGCCAGGCGAA 9731
777 spLupProMetAspAlaGluSerIleThrPheLysSerValSerAspLys 793
9732 GCGGCTCCGCGACACCTCTCTCCACGACGACCTCGCCCTATGATG 9781
794 AspLysArgGluLysAspLysProLysValLysSerAspLysThrLysAr 810
9782 TGCGCTATTTCGACGACACCGCGCCACTTCGACGACCGCGCGCT 9831
810 gLysSerAspGlySerAlaThrAlaLysLysAspAsnValLeuLysPro 827
9832 CTCTGCGCGCGACCGCAGC.....A 9854
827 eLysGlyProGlnGluLysValAspGlyAspArgGluLysSerProArg 843
9855 GCTCTCTCTCGCGCTCG.....ACTGCTGCGCCAGS 9886
844 SerGluProProLeuLys.....LysAlaLysGluGluAlaThrLysI 858
9887 ACAAGCCCGCGCGACACCGCTCTCGCGCGAGCGAAGCCAGC..... 9931
858 eAspSerValLysProSerSerSer.....SerGlnL 869
9932 .....GCAAGGCTCTCTCTCTCTCGGCGCAAGGCTCGCAGTGGA 9974
869 ys.Asp.....GluLysValThrGlyThrProArgLysAl 880
9975 AGGATGCGCCCTCTCCCTGCTGACCTCGCGGCTTCGCGGCTCAGC 10024
880 ahLysSerLysSerAlaLysAspThrArgArgLysSerGlnProArgThrA 897
10025 TCGAGCATGCGCAGCG.....CGCGCTCTCTCTCTCTCTCTCTCTCTCT 10065
897 rArgSerLysArgThrValProLysThrSerSerGlnLysSerGlnPro 913
10066 AGCGCTGCTCGCGCTCTCGCGCGCGAGCGCGCCCTCTCTCTCTCTCT 10115
914 ValArgThrArgArgProArgSerLeuArgLysIleAsnTyr..... 927
10116 CGTCGACGTCGTACAGCCGCTCTTTCGCGCATGCTCTCCCTGCGCG 10165
928 ....LeuIleAlaArgGluLysAsnGluArgGluLysArgLysLysSerV 943
10166 CCCTTGCGCGCTCGCTCGCGGTCGAGCCGCGCGCTCGCGCACAGC 10215
943 alAspLysAspPheGlu..SerSerMetLysIleSerLysValGluG 959
10216 CAGGGGAGATCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10255
959 LThrGluIleValLysProSerProLysArgLysMetGluGlyAspVal 975
10256 .....CCCTCGAGGACGCGGCGCATCGCGCCCTGCG 10288
976 GluLysLeuGluArgThrProGluLysAspLysIleAla.....Se 989
10289 GCAGGAAAGCGCTCACACCGTCGCGGCGCAAGCGGCGCATGCGCGCGCTC 10338
989 rSerThrThrPro.....AlaLysLysIleLysLeuAsnArgLysThrG 1004
10339 GAGCTGCGGCGCTCGACCTCGACCTACCTGCTCTCTGCGGCGACAGC 10388
1004 LysLys..... 1005
10389 GCCTCTCACCGCGCGCTCAACAGCCCGAGGCTACCTCTGTTCCGCGCG 10438
1006 .....LysI 1007

```

```

10439 AGCCCGCGCGCTCGACGCGCTGCTGACGCTCTCTCAACCGCCACCAAGT 10488
1007 eGlyAsnAlaGluAsnAlaSerThrThrLysGluProSerGluLysLeuG 1024
10489 TTGCGCGCGAAGATCGCGCTCGACTACG.....CTCCCATCTCGGCCCA 10532
1024 LysThrSerSerLysIleLysGlnGluLysValLysGlyLysAlaLys 1040
10533 GATGACAGCGCGCTCCAGACAGAGCTCGCCGACGTC..... 10567
1041 ArgLysValAlaGlySerGluLysSerSerThrLeuValAspTyrThr 1057
10568 .....TAGCCAAACATCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10593
1057 rSerThrSerSerThrGlyLysSerProValArgLysSerGluLysLysT 1074
10594 GAGCTCTCTCTTTATTCGACCGCTACCGCGACGACGAGCTCGACGCGCG 10643
1074 hrAspThrLysArgThrValIleLysThrMetGluGluTyrAsnAsnAsp 1090
10644 GCTCGACGCGCGCTACTGATATGAAACCTCGCGCAACCGCTCTGTTCT 10693
1091 AsnThrAlaProAlaGluAspValIleIleMetIleGluValProGlnSe 1107
10694 CGAGCGGACGAGCGCGCTCTCGACGATGGCATGCTCTCTCG..... 10738
1107 rLysThrAspLysAspAspPheGluSerGluGluGluAspValLysThr 1124
10739 .....TCGAGGT 10745
1124 hrGlnProIle.....GlnSerValGlyLysProSerSerIle 1136
10746 CAGCCCGCATCCGCTGCTACGCTCGCCCTCGCGGACCTCGGAGCGCT 10795
1137 IleLysAsnValThrThrLysProSerAlaThrAlaLysTyrThrGluY 1153
10796 CACCGCTCGATC.....CCGTCGTCGTCGCTCATTCAGACGAGAA 10836
1153 sGluSerGluGlnProGluLysLeuGlnLysLeuProLysGluAlaSerH 1170
10837 GAAGGCGACCTCGCC..... 10852
1170 IsGluLeuMetGlnIleGluLeuArgSerSerLysGlySerAlaSerSer 1186
10853 ....GCTGCTCTCTCTCTCTGCGGAGCTCTCTACCCGAGGCTGCGCG 10897
1187 GluLysGlyArgAlaLysAspArg..... 1194
10898 TCGACTGGAAGACTTCTTCGCGCCTACGCTCCGCGCAAGGTCTCCCTC 10947
1195 .....GluHisSerGlySerGluLysAspAsnPro... 1204
10948 CCCACCTACCCCTTCAGAGAGAGCGGTTCTGCTGACGCTCTCCAGGA 10997
1205 .....AspLysArg 1207
10998 CGAAGCTTCGCGACGTCGCTCCGCGAGGCTGACCTCGCGCGCAATTC 11047
1208 LysSerGlyAlaGlnProAspLysGluSerThrValAspArgLeuSerG 1224
11048 CGCTGCTCGCGCGCGCGCTCTTCGCGCAGCGCGGTGCTTCTCTCTT 11097
1224 uGlnGlyHisPheLysThrLeuSerGlnSerSerLysGluThrArgThrS 1241
11098 ACAGGCGCGCTCTCTCTCGAGAGCACCGCTGCTCGAAGCGCATCGCT 11147
1241 eGluLysHisGluSerValArgGly.....SerSerAsn..... 1252
11148 CTTCGCGACACCATCTACCGGCGACCGGCTTCTTCGACTGCGCCCTCG 11197
1253 .....LysAspPheThrProGlyArgAspLysLysValAs 1264
11198 ACGTGCGCGCACCGCTCGGCTCGACACCGCTCGAAGAGCTCAGCG..... 11242

```

| | | |
|-------|---|-------|
| 1264 | PTyAspSerArgAspArgTyrSerSerSerLysArgGlu..... | 1276 |
| | | |
| 11243 |TCGAAGGCCCTCTGCTCTGCTCCATCGCAGACACCCGCTCTCT | 11285 |
| 1276 | | 1276 |
| 11286 | CCAGATGTCCTCCGTCGGGCGCCCGTCGACAGCGACGACAGAGGGCGCTCTCT | 11335 |
| 1277 |AspGluArg..... | 1279 |
| | | |
| 11336 | TCCATAGCCGACACAGAGGAGCGCGGCTTCAAGATGCGCCCTGGACCTGCCAC | 11385 |
| 1280 | ... GlyGluLeuAlaArgArgLys..... AspSerProAlaGlyL | 1293 |
| | | |
| 11386 | GCCACGGGCTCTCTCTGCGGGGCGACCCCATCCCTCTCGCGCATCTCCA | 11435 |
| 1293 | ysGluSerLeuSerGlyLndLysSerLysLeuArgGlu..... | 1306 |
| | | |
| 11436 | CGAGTGGCTCCCTCGAGTGGCATCCCGGTGGACCTCGAAGGCTCTACG | 11485 |
| 1307 | ArgAspLeuProLysGlyAlaGluSerLysSerAsn.....Se | 1321 |
| | | |
| 11486 | CACACCTGCGCCACACCTCGGGCTTGGCTACGCGCCCGAGTCCAGGGGCTTC | 11535 |
| 1321 | rSerProAlaGAspArgLysLysProHisAspHisLysAlaProGluT | 1338 |
| | | |
| 11536 | CGCTCGCTGTACACACCGCGGCGACGAGCTCTTGGCCGAAGCCACGCTCCC | 11585 |
| 1338 | hLysArgPro.....Cys | 1342 |
| | | |
| 11586 | GGAACGGCGCGAAAGATGCCCGCGGTTGCCCTCCACCTGGCGCTCG | 11635 |
| 1343 | GluGluThrLysProValAspLysAsnSerGlyLysGluArgGlyLysHis | 1359 |
| | | |
| 11636 | TCGACAGCGCCCTCGATCATCTGCGCTTGGAGAGACAGACAGAGGGACG | 11685 |
| 1359 | salAlaGluAlaArgAsnGlyLysGluSerSerLysAla..... | 1372 |
| | | |
| 11686 | GTCCGCTGCGCCCTCTCGGAGCGGAAATCTCGCTCGCTCCGTCGCGTGC | 11735 |
| 1373 |AsnCysHisValTyrLeuThrArgGluThrLeuProThr | 1385 |
| | | |
| 11736 | CACACACCTGGCGCGCTGCCCTTCCACCGCTCCCAAGGGTGAATCTCCGCT | 11785 |
| 1386 | ArgArgSerTyr.....LeuLeuGlyArgTyrPArgArg | 1396 |
| | | |
| 11786 | CGATGCTCTGGCGACACCGCGAGGTGACACCTCTTGGCTCGGTGCAAGCG | 11835 |
| 1396 | gAlaProSerSerArgAsnPro | 1403 |
| | | |
| 11836 | CTCGCCATGCGGACGACGCTCG | 11857 |

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 17:23:55 ; Search time 46.67 Seconds

(Without alignments)
2890.710 Million cell updates/sec

Title: US-09-811-045a-1

Perfect score: 7374

Sequence: 1 MMEVDPNNKGMALNTGKTY.....WRRSWLLGRRRRAPSSRNPS 1404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR-71:*

1: p1r1.*

2: p1r2.*

3: p1r3.*

4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 6138.5 | 83.2 | 1560 | T42727 | proliferation pote |
| 2 | 3500.5 | 47.5 | 948 | A57640 | retinoblastoma bin |
| 3 | 1305.5 | 17.7 | 529 | T50609 | hypothetical prote |
| 4 | 480.5 | 6.5 | 1877 | T21861 | hypothetical prote |
| 5 | 383 | 5.2 | 3488 | T34418 | hypothetical prote |
| 6 | 374.5 | 5.1 | 5327 | T213564 | microtubule-associ |
| 7 | 363.5 | 4.9 | 6642 | T29757 | protein UNC-89 - C |
| 8 | 349 | 4.7 | 990 | T51618 | nucleolar phosphop |
| 9 | 342 | 4.6 | 1507 | B47328 | natural killer cel |
| 10 | 341 | 4.6 | 3968 | A44265 | trithorax homolog |
| 11 | 340 | 4.6 | 2738 | E88320 | protein F07A11.6 [|
| 12 | 339 | 4.6 | 2722 | T20532 | hypothetical prote |
| 13 | 337.5 | 4.6 | 1020 | QFHHH | neurofilament trip |
| 14 | 336 | 4.6 | 2526 | T20531 | hypothetical prote |
| 15 | 327 | 4.4 | 2187 | T30826 | nascent polypeptid |
| 16 | 326.5 | 4.4 | 1390 | S51364 | sperm tail-specifi |
| 17 | 325 | 4.4 | 1526 | A45605 | mature-parasite-in |
| 18 | 324 | 4.4 | 971 | T19431 | hypothetical prote |
| 19 | 321 | 4.4 | 2464 | 1 QRMSP1 | microtubule-associ |
| 20 | 320 | 4.3 | 1165 | T16420 | hypothetical prote |
| 21 | 320 | 4.3 | 1280 | T00365 | hypothetical prote |
| 22 | 316.5 | 4.3 | 606 | A43427 | neurofilament trip |
| 23 | 315 | 4.3 | 3869 | A48205 | All-1 protein +GPE |
| 24 | 314 | 4.3 | 1173 | T31421 | C-terminal domain- |
| 25 | 313.5 | 4.3 | 2649 | T51023 | hypothetical prote |
| 26 | 313 | 4.2 | 1274 | T16251 | hypothetical prote |
| 27 | 310 | 4.2 | 1791 | T02345 | hypothetical prote |
| 28 | 304.5 | 4.1 | 852 | T06310 | hypothetical prote |
| 29 | 302.5 | 4.1 | 1200 | A46194 | neurofilament prot |

| | | | | | | |
|----|-------|-----|------|---|--------|---------------------|
| 30 | 302 | 4.1 | 5170 | 2 | T15348 | hypothetical prote |
| 31 | 299.5 | 4.1 | 2364 | 2 | A56577 | microtubule-associ |
| 32 | 298.5 | 4.0 | 1390 | 2 | T14004 | trifA protein - Sli |
| 33 | 297 | 4.0 | 1192 | 2 | A71623 | probable secreted |
| 34 | 296.5 | 4.0 | 3147 | 2 | T18674 | hypothetical prote |
| 35 | 296.5 | 4.0 | 3938 | 2 | T42761 | Bassoon protein - |
| 36 | 296 | 4.0 | 1320 | 2 | J05630 | trCOF1 protein - mo |
| 37 | 295 | 4.0 | 749 | 2 | A45294 | Babini ring 2.1 |
| 38 | 294.5 | 4.0 | 1641 | 2 | T18614 | helicase II - huma |
| 39 | 294 | 4.0 | 699 | 2 | T38073 | nucleolar phosphop |
| 40 | 294 | 4.0 | 1403 | 1 | A47328 | natural killer cel |
| 41 | 293.5 | 4.0 | 990 | 2 | H86293 | protein T24D18.4 [|
| 42 | 290 | 3.9 | 1432 | 2 | B85431 | trichohyalin like |
| 43 | 288.5 | 3.9 | 2845 | 2 | T49505 | adenomatous polyo |
| 44 | 288 | 3.9 | 1870 | 2 | S37671 | MHC class III hist |
| 45 | 288 | 3.9 | 1871 | 2 | D96796 | probable heat shoc |

ALIGNMENTS

| | | |
|---|-------|---|
| RESULT | 1 | |
| T42727 | | |
| proliferation potential-related protein - mouse | | |
| C:Species: Mus musculus (house mouse) | | |
| C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Sep-2000 | | |
| C:Accession: T42727 | | |
| R:Witte, M.M.; Scott, R.E. | | |
| submitted to the EMBL Data Library, November 1998 | | |
| A:Reference number: Z22246 | | |
| A:Accession: T42727 | | |
| A:Status: preliminary; translated from GB/EMBL/DBJ | | |
| A:Molecule type: mRNA | | |
| A:Residues: 1-1560 <WTT> | | |
| A:Cross-references: EMBL:U83913; NID:9385884; PID:9385885; PIDN:AAC72432.1 | | |
| A:Experimental source: strain Balb/C | | |
| C:Genetics: | | |
| A:Gene: P2P-R | | |
| A:Function: | | |
| A:Description: involved in hnRNP association and Rb1 binding | | |
| A:Superfamily: RING finger homology | | |
| F:57-107/Domain: RING finger homology <RRN> | | |
| Query Match | 83.2% | Score 6138.5; DB 2; Length 1560; |
| Best Local Similarity | 85.4% | Pred. No. 3.1e-230; |
| Matches 1207; Conservative 36; Mismatches 133; Indels 37; Gaps 7; | | |
| QY | 1 | MMEVDPNNKGMALNTGKYATPTIDAAYATGKKKPPFLPEPPSSSEEDPIPAELL 60 |
| DB | 1 | MMEVDPNNKGMALNTGKYATPTIDAAYATGKKKPPFLPEPPSSSEEDPIPAELL 60 |
| QY | 61 | CLICDIMTDAVAVIPCCGSSSCDECIRTLLESKRTCTCHQNDVSPDALIANFLROA 120 |
| DB | 61 | CLICDIMTDAVAVIPCCGSSSCDECIRTLLESDEHTCTCHQNDVSPDALIANFLGRA 120 |
| QY | 121 | VNNFKNEGYTTRLRKQLPFLFLVPPRPPLSQRNLOPSSRSRPILRQDPVVFRTVST 180 |
| DB | 121 | VNNFKNEGYTTRLRKQLPFLFLVPPRPPLSMQRLQPLMRSPISNQDPLMIPTSSSA 180 |
| QY | 181 | CSDTTAGSCSDSGTLSPAPSSISLTSNOSSLAPVSGNSSAPAPVDDTATATVSI 240 |
| DB | 181 | HS-----APSSISLTSNPALAPSVSGNSSAPAPVDDTATATVSI 222 |
| QY | 241 | HSEKSDGPPRSDNKLPLPAALITSEHSKGASSIATLAMEEG---VPGTSPMNSIFVG 296 |
| DB | 223 | HSEKSDGPPRSDNKLPLPAALITSEHSKGASSIATLAMEEGGYQVPLGTBS---LLG 278 |
| QY | 297 | QSLRHQQLIPTGPVRINARPGGPGGHEHSHNKGILVSPQQIRRGRCYRSTINRR 356 |
| DB | 279 | QSLRHQQLIPTGPVRINARPGGPGGHEHSHNKGILVSPQIRRGRCYRSTINRR 338 |
| QY | 357 | HHSERQRTQSPSLPATPCFVVPVPPPLXPVPPHLLPLPGVVPQFSPQFSSQPTTAG 416 |

```
||||| 339 HHSRSQRTGSPSLPAPVFPVPPPLPPPHLPPLPGVLPQPSQPPQPPAG 398
417 YSVPPGFPAPANISTACFSPGYPTAHSNTPTQAPLISREEFYRQNDKGESEKFPY 476
399 YSVPPGFPAPANISTACFSPGYPTAHSNTPTQAPLISREEFYRQNDKGESEKFPY 458
477 GSGSYSSNSTYDS-SQGLAQIHALLTSPSAHTLIDLHDPH-PEEAARSAMIV 531
459 GSGSYSSNSTYDSKRSQSTRSRYSRFSRSHSRYSRSPRYRGRGKSRNRSRS 518
532 HMPDLMDIAHARSPPRRRRSRSPPEFRGSPPTKRNYPREKEKREYRNRYREVP 591
519 H-----GYHSRSRSPPTKRYHHSRSRSPQAFRGOSPTKRNYPRTGETEYFNRYREVP 573
592 YDIKAYGRSVDFRDPEKEKREYRMEKRYRMEYKYKGYAVGAQPPRSANREDFSPRL 651
574 YDIKAYGRSVDFRDPEKEKREYRMEKRYRMEYKYKGYAVGAQPPRSANREDFSPRL 633
652 LPLNIRNSPPTGRGRREDYAGQSHRNRLGNYPEKLISTRDSHNAKDNPKSESESNV 711
634 LPLNIRNSPPTGRGRREDYAGQSHRNRLGNYPEKLISTRDSHNAKDNPKSESESNV 693
712 GDGKGNHKKRRKRRNEKGESEFNLPELLETSRKCRGSSGIDETDTLFLVPSKDD 771
694 GDGKGNHKKRRKRRNEKGESEFNLPELLETSRKCRGSSGIDETDTLFLVPSKDD 753
772 ATPPRDEPMDESIITFKSVSDKREKDPVKVSDKTRKDGSAVLAKDVLVPSKQP 831
754 ATPPRDEPMDESIITFKSVSDKREKDPVKVSDKTRKDGSAVLAKDVLVPSKQP 813
832 EKVVDREKSRSEBPLKAKAEATKIDSVKSSSQDEKVTGTPRKASHKADTRRQ 891
814 EKVVDREKSRSEBPLKAKAEATKIDSVKSSSQDEKVTGTPRKASHKADTRRQ 873
892 SQPTRRSKRTVPTSSQSPVTRRRSLKTNLYLAR-EKDEREKAKSVNDKESS 950
874 KPADEKRYKXCSKDISEKSPASDEKAKKPEKKLLDSKEKKRRKKEESVNDKESS 933
951 SMKISKVGEITVPSPRKKEGDEVEKLEPTPEKDKIASSTPAKKITLNETKTKIGNA 1010
934 SMKISKVGEITVPSPRKKEGDEVEKLEPTPEKDKIASSTPAKKITLNETKTKIGNA 993
1011 ENASTTKEPSEKLESTSSKIKOEKVKGKAKRVAGSESSSTLVDTYSTSGSPVAKS 1070
994 ENASTTKEPSEKLESTSSKIKOEKVKGKAKRVAGSESSSTLVDTYSTSGSPVAKS 1053
1071 EEKDTKRTVKTMEEYNNNDTAPAEVYIMIQVPQSKWDDDFESEBEDVKTQPIQSV 1130
1054 EEKDTKRTVKTMEEYNNNDTAPAEVYIMIHVPQSKWDDDFESEBEDVKTQPIQSV 1113
1131 GKPSIITKNVTTPSATAKYTEKESEDOPEKLOKLPKEASHELMOHELSSGSSASSEKGR 1190
1114 GKPSIITKNVTTPSATAKYTEKESEDOPEKLOKLPKEASHELMOHELSSGSSASSEKGR 1173
1191 AKDBHSGSEKDNPKRRSGAQPDKESTVDRLSEQGHFKTLQSQSKETRTSEKHESVAGS 1250
1174 AKDBHSGSEKDNPKRRSGAQPDKESTVDRLSEQGHFKTLQSQSKETRTSEKHESVAGS 1233
1251 SNKFTTGRDKKVDYDSRDYSSSKRDERGLARKDSPPRGKESLSQOKKLEEDRLP 1310
1234 SNKFTTGRDKKVDYDSRDYSSSKRDERGLARKDSPPRGKESLSQOKKLEEDRLP 1293
1311 KKGAEKSSNSPPRDKKPHBHKAPYETKRPCEETKPPDKKSGKEREHAHAARNGESS 1370
1294 KKGAEKSSNSPPRDKKPHBHKAPYETKRPCEETKPPDKKSGKEREHAHAARNGESS 1353
1371 GANCHVYLTRQTLFWRBSWLLGRWRAPSSSRNP 1403
1354 GGRKLPCLINPDLMEKELAVGVEKSAVVKPKP 1386
```

```
A57640
retinoblastoma binding protein RBO-1 - human
C:Species: Homo sapiens (man)
C:Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 17-Mar-2000
C:Accession: A57640
R:Sakai, Y.; Saijo, M.; Coelho, K.; Kishino, T.; Nikiwa, N.; Taya, Y.
Genomics 30, 98-101, 1995
A:Title: cDNA sequence and chromosomal localization of a novel human protein, RBO-1 (
A:Reference number: A57640; MIM:6129310
A:Accession: A57640
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-948 <SAK>
A:Cross-references: GB:A85133; NID:q728590; PIDN:CA59445.1; PID:q755748
C:Genetics:
A:Gene: GDB:RBBP6
A:Cross-references: GDB:626076
A:Map position: 16p12-16p11.2
C:Superfamily: RING finger homology
C:Keywords: tandem repeat; zinc
F:79-129/Domain: RING finger homology <RNG>

Query Match 47.58; Score 3500.5; DB 2; Length 948;
Best Local Similarity 73.28; Pred. No. 1,9e-128;
Matches 691; Conservative 44; Mismatches 130; Indels 79; Gaps 10;

QY 1 MMEYKDPNMKGAMLTNNGKVAIPTDAEAYAGKKEKPPFLPEEPPSSSEEDDPIPAELL 60
DB 23 MMEYKDPNMKGAMLTNNGKVAIPTDAEAYAGKKEKPPFLPEEPPSSSEEDDPIPAELL 82
QY 61 CLICDINTDAVVIPICCGNSCDECIPTTLLESDKHTCPTCHQNDVSPDALIANKFLRQA 120
DB 83 CLICDINTDAVVIPICCGNSCDECIPTTLLESDKHTCPTCHQNDVSPDALIANKFLRQA 142
QY 121 VNNFNKNETGYTKRLRKQPLFLVPPRPLSQNLDPKRSNPLTRQODPVVFYTYTSPT 180
DB 143 VNNFNKNETGYTKRLRKQPLFLVPPRPLSQNLDPKRSNPLTRQODPVVFYTYTSPT 202
QY 181 CSDDTKTAGSCSDSGTLRLPAPSSISLTSSNOSLAPVSGNPSAPAPVDITATVTSISV 240
DB 203 -----HAPSSISLTSSNOSLAPVSGNPSAPAPVDITATVTSISV 244
QY 241 HSEKSDGPFRRSDNKLPAALATSEHSKASSIATLAMEBK-----VPGTSPWNSIFVG 296
DB 245 HSEKSDGPFRRSDNKLPAALATSEHSKASSIATLAMEBKGYQVYVLGTPS-----LLG 300
QY 297 QSLHGGILPTTGVRTNARPGGGRGWEHSNKLGYLVSPQOIRRGESCYRSINRGR 356
DB 301 QSLHGGILPTTGVRTNARPGGGRGWEHSNKLGYLVSPQOIRRGESCYRSINRGR 360
QY 357 HHSRSQRTGSPSLPAPVFPVPPPLPPPHLPPLPGVLPQPSQPPQPPAG 416
DB 361 HHSRSQRTGSPSLPAPVFPVPPPLPPPHLPPLPGVLPQPSQPPQPPAG 420
QY 417 YSVPPGFPAPANISTACFSPGYPTAHSNTPTQAPLISREEFYREQ----- 465
DB 421 YSVPPGFPAPANISTACFSPGYPTAHSNTPTQAPLISREEFYREQ----- 480
QY 466 -----NDKRE-----SKFYSSGSSYSRSTYDS-SQGLAQIHALLT 501
DB 481 KLDEFNDFAKELMEYKIKERRRSFSRSPYSSGSSYSRSTYDS-SQGLAQIHALLT 540
QY 502 LSPSAHTLIDLHDPH-PEEAARSAMIVHMDLMDIAHARSPPRYRRRSR 557
DB 541 RSPRSRSTYDSRSPRYRRRSRSH-----GYHSRSRSPRYRRRSR 595
QY 558 SPPEFRGOSPTKRNVPREEKREYFNRYREVPYDIKAYGRSVDFRDPEKEKREYME 617
DB 596 SPQAFRGOSPTKRNVPREEKREYFNRYREVPYDIKAYGRSVDFRDPEKEKREYME 655
QY 618 RKYREYKYYKGYAVGAQPPRSANREDFSPRLPLNIRNSPPTGRGRREDYAGQSHRN 677
```

Db 656 RRYREWEKYYKGYAAGQPRPSANRENPSPERFLPLNIRNSPTFRGRREDYVGQSHRS 715
QY 678 RNLGNTYEPEKISTRDSDHNAKDNPKSKESESENVPGDGNKKHKKRRRNEEGESSSEF 737
Db 716 RNLGNTYEPEKISTRDSDHNAKDNPKSKESESENVPGDGNKKHKKRRRNEEGESSSEF 772
QY 738 LNPLELLETSRKCRSGSSGIDETDITFLVPSRDATPVPRDEMDSEITFEKSVSDKDR 797
Db 773 LNPLELLETSRKCRSGSSGIDETDITFLVPSRDATPVPRDEMDSEITFEKSVSDKDR 832
QY 798 KDKPRKSDKTRKRSKSDGATKADNVLPKSGQPEKVDGDEKSPRSEPLTKAKEEATK 857
Db 833 RDKPRKAGDKTRKKNMGSAVSKENIVAKPAQPOKEKVDGDRDLDMQLKKPRRLRLR 892
QY 858 IDSKVPSSSSOKDEKVTGTPPKAHKSAKADTRROSQPTRRNSKR 901
Db 893 LTLINHLPLRRMKKSLPEPEKL-----TLNQOKTPRRNKTSQR 930

RESULT 3
T50609
hypothetical protein DKFZp761B2423.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50609
R:Bioecker, H.; Bioecker, M.; Brandt, P.; Mewes, H.W.; Well, B.; Wiemann, S.
submitted to the Protein Sequence Database, June 2000
A:Reference number: 225143
A:Accession: T50609
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-529 <AAA>
A:Cross-references: EMBL:AL359564
A:Experimental source: adult amygdala; clone DKFZp761B2423
C:Genetics:
A:Note: DKFZp761B2423.1

Query Match 17.7%; Score 1305.5; DB 2; Length 529;
Best Local Similarity 77.9%; Pred. No. 7.5e-44;
Matches 250; Conservative 32; Mismatches 38; Indels 1; Gaps 1;
QY 1052 TLVDYTTSTSTGSPVPRKSEKTDTRKTYIKTMEYNNNDNTPADVDIIMIOVPSKMDK 1111
Db 1 TLVDYTTSTSTGSPVPRKSEKTDTRKTYIKTMEYNNNDNTPADVDIIMIOVPSKMDK 60
QY 1112 DDFESEEDVKTQPIQGVKPSIIKNVTKPSATARYTEKESQPEKIQKLPKEASHE 1171
Db 61 DDFESEEDVKTQPIQGVKPSIIKNVTKPSATARYTEKESQPEKIQKLPKEASHE 120
QY 1172 LMOHLEBSKSGASSEKGRADREHSGSEKNDPKRKSQAOPDKESTYDRLSEOGHFTL 1231
Db 121 LMOHLEBSKSGASSEKGRADREHSGSEKNDPKRKSQAOPDKESTYDRLSEOGHFTL 180
QY 1232 SOSSEKTRTSEKHEVNRSSNKFDPGRDKKYDSDVSDVSSSKRRDERGELARRDSPPR 1291
Db 181 SOSSEKTRTSEKHEVNRSSNKFDPGRDKKYDSDVSDVSSSKRRDERGELARRDSPPR 240
QY 1292 GKEISLGQSKLREBRDLPPKKG-AESKSSNSSPRPDKRPHDKAPYETRRPCEETPYDK 1350
Db 241 NKDSASGQKNKPREERDLPKKGTDGSKSSNSSPRDKRPHDKAPYETRRPCEETPYDK 300
QY 1351 NSGKERKHAALFARNGESSG 1371
Db 301 NSGKERKHAALFARNGESSG 321

RESULT 4
T21861
hypothetical protein F36F2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21861

R:Cottage, A.
submitted to the EMBL Data Library, November 1996
A:Reference number: 219479
A:Accession: T21861
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1877 <WTL>
A:Cross-references: EMBL:Z81532; PIDN:CA804326.1; GSPDB:GN00019; CESP:F36F2.3
A:Experimental source: clone F36F2
C:Genetics:
A:Gene: CESP:F36F2.3
A:Map position: 1
A:introns: 49/3; 86/2; 112/3; 139/2; 235/3; 284/3; 436/3; 507/2; 566/2; 678/2; 1000/3

Query Match 6.5%; Score 480.5; DB 2; Length 1877;
Best Local Similarity 19.0%; Pred. No. 2.2e-11;
Matches 276; Conservative 168; Mismatches 400; Indels 605; Gaps 51;
QY 1 MME--VKDPNMGAULTNKGKAIPTI--DAEVAIGKKKKRPFLEPSESSSEEDDPIR 56
Db 249 IMETVDDPD--AAHBSGKYVIFIMHWKARQETLARKN-----DGSSPAQTSRKVP 300
QY 57 AELLCICKDITDAVVIIPCCGNSCDECIPTTLESDEKHTCP--TCHONDVSPALIAN 114
Db 301 PELLCIPICSLKEAIVISCCGNSYCADCIEARILDPDQKCPGADCGK-DISTSIIPN 359
QY 115 KFLQAVNNKNEGTYYKRLKQDLPFLVLPVPPPLSGRLQDPSRSPILRODPVVR 174
Db 360 KTLRDAAMAMLSATGPAFTTPQIVP-----BPEQIR 391
QY 175 YTVSPCTGDTKAGSCSDGTLISRLPAPSISSLTNGSSLAPVSGNPSAPAPVDITA 234
Db 392 IRIG-----LKAPS-----SSQSIITP----- 408
QY 235 TVSISVHSEKSDGPPRSDNKLPAALVTEHSGKASSIAITALEMEKGPSTSPWNSIF 294
Db 409 -----SGISPGSTLV 418
QY 295 VGOSILHGLITPTGPVRINARPGGPRGWEHNSKLGVLSPQOIRRGERSCYRINR 354
Db 419 OOOTL-----TSVS 429
QY 355 GHHHSERQRTOSPLRATPCFPVP---PPLYPPTPTLPLPGVPP--PQSPQFSS 410
Db 430 GTSLSAQPSNV-NPSILGIPILASQVSMQVDSLPPQLQDELPPGILPQF----- 481
QY 411 GPPTAGVSVPFGFP-----PAPANISTACFSPGVPTAHSNTPTTQAPLLSRE 459
Db 482 -----GLPPGCVGLSATVLPQHOQSMPLNYGMPLFSAGFPAYS-----SVPRSAISD 530
QY 460 EFYRQNDKGRSKRPYSSGYSRSSYTSOGLAQHIALTLSPAAHTLDDLHDHPH 519
Db 531 EWNAAFQNDKRS-----SRDHKDRTRKRKH----- 558
QY 520 PEELARSAAMYHMDLMDIAHARSPPYRRTYRSRSRSPPEPRGOSPTKRNVPREKER 579
Db 559 -----DSRSR-----RRDSSSSSSSMSSSSSDEDERRRKREKES 593
QY 580 EYFNRYREVPPYDIKAYYGRSVDFDPPEKERYEWEKREWEKYYKGYAAGQAPR 639
Db 594 SKKRSVKEKP-----RRADERRDRDRDRDRDRSHD----- 629
QY 640 SANREDSPERLLPLNIRNSPPTGRREDYAGGSHRNRLGNTPEKLTSTRDSHNAKDN 699
Db 630 -----VRS-----GSKDKIKASSHR-----RD-----RD 651
QY 700 PKSKESSENVPGDGNKKHKKRRRNEEGESSSEFLNPELLETSRKCRSGSSGIDETK 759
Db 652 ARKDRRRDDV-----RKEREKREED-----DDOK 679
QY 760 TPTLFLVPSRDATPVPRDEMDSEITFEKSVSDKDRKREKDKPVKSDTKRKSQSGSATK 819

C:Species: Drosophila melanogaster
 C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13564
 R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
 Submitted to the EMBL Data Library, April 1999
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
 A:Reference number: 217689
 A:Accession: T13564
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-5327 <SPA>
 A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
 C:Genetics:
 A:Cross-references: FLYBase:FBgn0025392
 A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
 A:Note: EG:4984.1
 C:Superfamily: Drosophila 576k microtubule-associated protein homolog

Query Match 5.1%; Score 374.5; DB 2; Length 5327;
 Best Local Similarity 21.5%; Pred. No. 8.2e-07;
 Matches 317; Conservative 217; Mismatches 606; Indels 333; Gaps 69;

QY 19 KYAFTIDAENYAIKKKKPPFLPEEPSSSEED--DPIPAELLCLICKDITDAV--- 72
 DB 983 KAEVODIATFAKKDIKSR---TEEQALAKPAEELSSPTPEKLSKTSQDKKDOIGAPV 1038
 QY 73 -VIPCCGNSCDECIKRTTLLSEDKHCTPTQOND-----VSPDALLANKELQAVNNFK 125
 DB 1039 DVLPVNLQESLPEKFKFTSIESGATTAPTLPEDERIPLDQIKEDLVIEEKYKEE----- 1093
 QY 126 NENGYTRLRKQPLFLVPPRPRLSQRNLQPPRSRSPILNQDPVYVRYVYPTCSDTK 185
 DB 1094 -----TEEAIVAVVQTLPEAAPLAIDTI-----LASATK 1125
 QY 186 TAGSCSDSGTSLRPL-----APSIISLTSSNOSSLAPVSGNPSSAPAPVPDITATVSI 239
 DB 1126 DAPDANNAEALGELPDGSEYRLPKMFTFEAQONLLRDYK-----TDEVA--DLP 1174
 QY 240 VHSKSDGP--RQSDNKLPPAALTSEHSKGASSIATALMEKGV-----G 286
 DB 1175 VHEADLGLYEKDSQD-----ANAKSISHEKSA-----EKEETDEKENKVEIELG 1223
 QY 287 TSPNSISFVGGSLHGO-----LIPTGPRINARGGGPRGHEHKNKLGVLSP 337
 DB 1224 DEP--NKVDISHVLKESVOEAEKVAVVETVEKKOEIVETVITQENQEDLMEOVKD 1282
 QY 338 ----POQIRRG--ERSCYRSINRGHHSERSQRTQSPSLPATPCFVVPPLVPPPH 390
 DB 1283 KEHEOKTESGIIITEKEKKKAST--PEEKETSDITSDELPA----- 1323
 QY 391 TLPLPGVPPQPSPPQPS--SOPPT-----AGYSVPP-----GFP 426
 DB 1324 QLADPTVPPKASADREDTGSIESPRTIEALEVEVQAQQAQKVPVPAEAIKTEKSP 1383
 QY 427 APAIISTACFPGVPTAHSNMPTTQAPLSREFFYRQONKGRSKPPYSGSSISR--SS 485
 DB 1384 ASKETSRPESYTGSKVEDTEQTKSKSPVSRPE-----SEADDKKSPFASGEASRPES 1437
 QY 486 YTDSOGIAGHIIHALTISPSAAH---TLDLLHHPPEAEARSAMIVHMDLMDIAH 541
 DB 1438 VAEVQKQMGKAEBSRESIATHKDESLD-----KAKQESR-----RESIAESIK 1484
 QY 542 ARSSSPYRRYRSRSPPEF--RGOSPTKRVNPREKEBEYFNRYRVPPYDIKAYV 598
 DB 1485 PESGIDEKSALESKPEASVTDKSKPESRESIAESILKAESTDKESAPPSKASRP 1544
 QY 599 GRVDFRPFPEKERYREMYEKYKGYAVGAQPPRSANREDSPPERLPLNLRN 658
 DB 1545 SVESVVKD--LTEKSKESR-----ESIAESAP--PIEFREVSRRPSVID----- 1587
 QY 659 SPFTGRREDYAAQSHNRNLG--GNYPKLTSTRDSHNAKNP--KKEK--ESENVPGD 713

DB 1588 -----GIKDESAKPESSRRDSESLAKESRPESV-----LESYKDEPIKSTEKSRRESVAES 1638
 QY 714 GKGNKKHKKRRRRNEKEGESESEFLNPBLTSSKRCRSSGIDETKDTLVLPSRQAT 773
 DB 1639 FKAOSTK-----DEKSPLTCKDISRPE-----SAV--EWMVAPKETSRRPSA 1680
 QY 774 --PYRDEPMAESITFKVSD--KDKREKDKPKYSDKTKRKSQGSATAKDNVLPKSG 829
 DB 1681 VGSMDKDESMKPEPSRREYVKGGAQSRISRPASVAESA---DQADLKE--LSRPEST 1735
 QY 830 POEYVGD--REKSPREPLPKAKEATKIDSVPSSSQKDEKYGTGTPKRAHSAKD 887
 DB 1736 TQSKREAGSIKDEKSP-----LASEEASRPASV--AESYKDEA-----EKSKEE 1776
 QY 888 TRROSQPTTRSKRTVPTSSQKQPVTRRPRSLRKNYLLAREKNREKRXSYNDKF 947
 DB 1777 SRRESVAEKS-----PLPSKEASRPASV--AESIKD-----EAKSKESERRRESVAEKS 1823
 QY 948 ESSSMKISKVEGT--EIVKPSKRRMEGDVEKL--ERTPEKDIAS--STTPAKTK--L 999
 DB 1824 PLPSKEASRPASVAESIKDEAEKSEBSRRESVAEKSPLPSKEASRPASVAESIKDEAEK 1883
 QY 1000 NRETGKRIAGNAENASTKEPSEKLESTSSKIKQEKVKGAK--RKYVAGSEGSSTLVDTYS 1058
 DB 1884 SKESRRESVAEKSPLPSKEASRPASVAESIKDEAEKSEBSRRESVAEKSPLPSKEASR 1943
 QY 1059 TSSYGGSPVRSKSEKTDPTKRVITKMEYNNDNTPAADVLIIMIOVPS---KMDKDFE 1115
 DB 1944 PASVAESIKDEAEKSEKESR-----RESVAEKSPLPSKEASRPASVAESIKDEAEKSE 1998
 QY 1116 SEEDVY--KTQPIQSVKPSITIKVNTKPSAVAKYKESQEQEKLOKLP--KEASH 1170
 DB 1999 SRRESVAEKSPLPSKEASRPASVAESIKDE-----AEKSKESRRESVAEKSPLPSKEKSR 2054
 QY 1171 -----ELMOHELRSK-----GSASSEKGRADREHS-----GSEKDPNDRKSGAQDPK 1215
 DB 2055 PASVAESIKDEAEKSEKSRRESVAEKSPLPSKEASRPASVAESIKDEADSK-----E 2108
 QY 1216 ESTYDRLSEQGHFKTL---SOSKETRTSEKHESVROSSKNDFTPGDKKVDYDSRYS 1271
 DB 2109 ESRRESVAESKQAOSIKQDOSPLKEVSPRESVAESVKDDPKSKPSSRESV---AGSVT 2165
 QY 1272 SSKRDERGELARRKDDSPRCKESLSGQSKLREBRDLPKGAESKFNSSPPRDKKPHD 1331
 DB 2166 ADSARDQSPLESKASRP---ESV---VDSYKDEAEQESRRESKTESVILPPAK---D 2216
 QY 1332 HKAPYETKRPCEETKPPVDKNSGKEREKHAABAR 1364
 DB 2217 DKSPREVLQPVSMETIREDADQPKKPSQAESR 2249

RESULT 7
 729757
 protein UNC-89 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
 C:Accession: T29757
 R:Du, Z.; Le, T.T.; Wilson, R.
 submitted to the EMBL Data Library, May 1997
 A:Description: The sequence of C. elegans cosmid C09D1.
 A:Reference number: 220679
 A:Accession: T29757
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-6642 <DUZ>
 A:Cross-references: EMBL:AF003131; PIDN:AA54132.1; GSPDB:GN00019; CESP:unc-89
 A:Experimental source: strain Bristol N2; clone C09D1
 C:Genetics:
 A:Gene: CESP:unc-89
 A:Map position: 1
 A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1
 /3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

| | | | | |
|-----------------------|-------------------|--------------------|-------------|--------------|
| Query Match | 4.9%; | Score 363.5; | DB 2; | Length 6642; |
| Best Local Similarity | 24.5%; | Pred. No. 2.8e-06; | | |
| Matches 180; | Conservative 120; | Mismatches 260; | Indels 175; | Gaps 34 |

[illegible]

RESULT 8
151618
nucleolar phosphoprotein - African clawed frog
C:Species: *Xenopus laevis* (African clawed frog)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C:Accession: 151618; S57757
R:Calins, C.; McStay, B.
J. Cell Sci. 108, 3339-3347, 1995
A>Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180
A:Reference number: 151618; MUID:96019267
A:Accession: 151618
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-990 <CAB>
A:Cross-references: EMBL:X88927; NID:9895920; PIDN:CAAC1368.1; PID:9895921

C;Genetics:
A;Gene: xNOPp180
C;Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C;Keywords: phosphoprotein

| | | | | |
|-----------------------|-------|---|-----------------|------------|
| Query Match | 4.7% | Score 349 | DB 2 | Length 990 |
| Best Local Similarity | 22.5% | Pred. No.1,4e-06 | | |
| Matches | 201 | Conservative 114 | Missmatches 420 | Indels 156 |
| Gaps | | | | 33 |
| QY | 536 | LMDIAHARSPPYRYRARSRSRPPEFGQSPTRKRVVPREEKEREYFNRYREVPYDRIK | 595 | |
| Db | 47 | LDDIFSDWAKSPD-----AKKKRP---ANGLPKKSAAKSSSEDSSEDEDEPP----- | 93 | |
| QY | 596 | AYGRSVDFRDPPEKEREREMERKRYREWKYKGYAAGAPRPS-ANREDSPPELPL | 654 | |
| Db | 94 | -----AKKRAOPAGCK-----KPVAAVQPKAKASSSEDSSEDSDEE | 131 | |
| QY | 655 | NIRNSPFTB-GRREDDAAGOSHNRNLGNGVPEKLTSPDSHNAKONPKSKERESEVPPD | 713 | |
| Db | 132 | ETKPKPAKPAKPAQPPKYAAAYKTPQKAKAKSSSESSSEDEBAKKKQPVYIKVPPKAVV-- | 189 | |
| QY | 714 | GKGNKHKHKHRRKRNKEKGESSEFLNPELLETSSRKCRGSSGIDETKDTPLVLSRDAAT | 773 | |
| Db | 190 | -KAGLASNNGKATDSSSEDS | 229 | |
| QY | 774 | PVRDEPMDAESITFKSVSDKDKREKPKVKSXTKTRKD-----GSATAKKDWLAP | 826 | |
| Db | 230 | AAK--POAKRTAGKAKSSSRDSSDSEDEQKTAKSPPKDPVYSAVPPVSVSKKTLTSP | 287 | |
| QY | 827 | SKPOEKVGDREKSPREPEPLTKAKEEATKIDSVPSSSQKDEKVTGTPRKHAHSKAK | 886 | |
| Db | 288 | GTYKAKESDSSDSDSEDEBPAKKAKIVAKAASAAPKLAKAE--TSPDSESDSSSD | 345 | |
| QY | 887 | DYRKQSPTRNSKRTKVPPTSSOKSOPVTRPRSLKINYLARKNREKKKKSYDND | 946 | |
| Db | 346 | EKKSYSYKLGKAPKAPAPADAKSPV-----AAAKKSAAPKASSSD | 390 | |
| QY | 947 | FEKSSKAKISVEGTETLVKPSPK-RKMEGVDEKLETPPEKDKIA-SSTTPAKKIKLIRETG | 1004 | |
| Db | 391 | SDSSNE-----ETT--TKPAAKTTPAKSATPTPSKPTPNKATPTKTPAKPOTPKTSTA | 444 | |
| QY | 1005 | KIKGNAMNS-----TKEPSEK--LESTSSKIKOEKVGA--KRVAGSEGSS | 1051 | |
| Db | 445 | KRSSSSDSSDSEDETTTKPAAKTTPAKSATPTPSKPTPNKATPTSKTTPAKPGTAP | 504 | |
| QY | 1052 | TLDVYTSSTGSGPYRKSEKTDTRT-----VIKMEYNDNDAPADVIMT-1102 | | |
| Db | 505 | TSAAKRDSSDSSDSSDCKTPPAKTTPAKPAKTTPAKPAKTTPAKPAKSTP | 564 | |
| QY | 1103 | ---OVP-----OSKMDKDFESEED-----YKTPPIOSVGSPSSIIKNVTKRPSATAK | 1149 | |
| Db | 565 | GKQVPIPKKSSSSDSSDSEDEKKSAPKAVTMT--PGKATSKPVVASKRVPAKKASSS | 623 | |
| QY | 1150 | YTEKSEDEPKILOKLPKEASHELMOHELNASSKGSASSEKGRADREHSGSEKONPDKRS | 1209 | |
| Db | 624 | DSOSEETETTKTKPLTKLSPAV--KTLPPKAKSSSD-----SSSDSESEKTKP | 672 | |
| QY | 1210 | GAQPDKESTYDRLSEGHFTL-----SQSSKETPTSEKHESVRGSSNNDPFGDKVY | 1263 | |
| Db | 673 | AKPPASATPVNTKAPQONKASAKSCSDSSSSEEBGKSQOTPKGSPAATATPPKKNPV | 732 | |
| QY | 1264 | DYDSRQYSSSKRRDERGELARRDSDPPRGKESLSGOKSKLREERDLPKKAESKSNSSP | 1323 | |
| Db | 733 | AVAKDPRSSSSSSDSDGDEKOK--PKQAAAKDVQKQAKAKPPPKAASSSESSS | 789 | |
| QY | 1324 | PRD---KKPHD--HKAPYETKRCETKPVDNKSGKEREKHAALANGKSS | 1370 | |
| Db | 790 | DEDVSAKKTNTAVSKSPVYTPKAVPAK--KESSESSESDSEKOGKNTS | 839 | |
| RESULT | | | | |
| B47328 | | | | |
| | 9 | | | |

natural killer cell tumor-recognition protein - mouse
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tum
C:Species: Mus musculus (house mouse)
C>Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 05-Nov-1999
C:Accession: B47328; I77662
S:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A:Title: A cyclophilin-related protein involved in the function of natural killer cells.
A:Reference number: A47328; MUID:93133824
A:Accession: B47328
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1507 <AND>
A:Cross-references: GB:I04289; NID:g192866
A>Note: authors translated the codon AGT for residue 972 as Arg
R:Rinfret, A.; Anderson, S.K.
Mol. Immunol. 30, 1307-1313, 1993
A:Title: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicing m
A:Reference number: I57820; MUID:94019422
A:Accession: I77662
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 235-237;263-294 <RIN>
A:Cross-references: GB:S65998; NID:g425701; PIDN:AAB28500.1; PID:g425702
C:Genetics:
A:Gene: NK-TR
C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C:Keywords: alternative splicing; lymphocyte
F:60-230/Domain: cyclophilin homology <CYP>

Query Match 4.6%; Score 342; DB 2; Length 1507;
Best Local Similarity 20.1%; Pred. No. 4e-06;

Matches 304; Conservative 186; Mismatches 569; Indels 456; Gaps 65;

65 KDIMDAVYIPCCGNSCDECIPTLLESD---KHTCPTCHQ-----NDVSPD 109
Db EMLKTDASRPADRVIDCGVLATKIKDYEPKKRKRKPTCEGSDSSRSSSSSSSE 267
Qy 110 ALIANKFLRQAVNNFNKNETGYTKRLKQLPPLFLVPPRPLSQRLNP----- 158
Db SEVERETIRRRHRRKRRPKVRAHKKRRKREMS-----SSEPRKRRTVSPEGYSERSDVE 321
Qy 159 -RSRPDLRQODPVYFRTVSPCTDTAGSCSDSGLTSLRPAISIS-SLTSNOSSLAP 216
Db KRVSQNTKREKPVVPEIPVY-----PENRFLRDMPAITVEPQNIIPVAP 371
Qy 217 PVSGMPSAPAPVDPITATVSIVHS---EKSDGPFRODNKLLP-----AA 260
Db VSDQKPVSVKSGRKIKGRGTRIRYHTPPRSRSHSESKDDDSSETPPHKMEQRLRAYRP 431
Qy 261 ALTSHSKGASSIAITALMEKGVPTSPWNSIFVQ----- 297
Db PSGEKMSKG-----DKLSDPCSSRWDERSLQSRSWSYNGYSDLSSTARSHDGH 481
Qy 298 -----SLMGQLIPTGPPR-----INARPG 320
Db HKHHRKKFKHKKRKAOKHCRHRHQTKKRIIVMDLEPSSPHRMKSSCVRRERSRA 541
Qy 321 GRPGEHNSKLGVLVSPPOQ---IRGERSCYRSINRGHNHSESRORTO----- 366
Db SSSSHSHSKRMWSKSDODDGSASTHSSRDSYRSKSHSDBSGRSRAVAVKKSSRSIN 601
Qy 367 -----SPSLPATPCFVPPPPPLYP-PPHTLPLPGVP-----PPOFSPOFP-- 408
Db RSKSRSSSSSGPRTSISPKPAQLSENKPVTEPLRPSPONGVILVQVVAENIPVIR 661
Qy 409 -SSOPTAGYVPPPPPPAPANISTAGSPGVPTAHSTMTQTQAPLSREFFREQND 467
Db LQSDSPPSKWK---PGQKP-----WKPSYERIQEKAKATTN--LLPVOSTYSLNVI 707
Qy 468 KGRSKFPY-----SGSSYSRSSYDSSQGLAQHIAHTLSPSAHITDLIDHDPH 518

Db KATVSSSYHKKRKPESDGSAYSK--YSDRSSG----- 739
Qy 519 PPEEASAMVYHMPDLMDIAHARSPPPIKRRYSRSRSPPEFGQSTKRNVPREKE 578
Db 740 -----SSGRSGS-----KSSRSRSSRSYTSRSRS-----LPTSLSLSRSS 778
Qy 579 REYF-NRREVPYVDIKAYY-----GRSVDFRDPFEK-----ERYEMERKYNEWYEK 626
Db RSHSNKYSIDGQSHRSSSYTSVSSDDGRAMFRSNRKKSVTSKRRHNSNKK--TLASK 836
Qy 627 YKGYAVGAQPPRPSANRDEPSPERLLPLINRSPPTRGARE-DYA--AGQSH----- 675
Db 837 YVRG-----REKSS-----RRKXSSSSSLDYTSDDQSHVQYYSAP 875
Qy 676 -----RNRNLGNYPE-----KLSTRSHNAKNKPKSKKESEVPGDG 714
Db KEKQKVEALNDKQKGRGKPEKPEWECPRSKKENSEDH--SRDVSFKGKKACAGSKMDS 934
Qy 715 KGN-----KHKHKKRRHNEKGE--ESPSFLNPELLETSRKCRGSGSIDETKTDTFLVLP 767
Db 935 ESNSEQDYTKSKSPDKRKGSEKEGEASDSESEVGQSHITAK-----PPAKPTSTPLP 989
Qy 768 SRDDATPVDEPMDAESITFKSVS-----DKDRKDKPKVSKDKTKRKSDGSATAK 820
Db GSDGAMKSR-RQSSASSESSCSNLGNIRGEPQKQKSHKDLKGDHTRKAREKS-KAKK 1047
Qy 821 DNVLKPSKQPOKVDGDRKSPRSEPPPLKAKAEATKIDSVKSSSSOKDEKVTGTP-K 879
Db 1048 D---KHKRKP-----RKQAFHMQPLEFGDDEEEMNG-----KQVTDQDPK 1088
Qy 880 ASKSAKQTRKQSQPRTKRSKTYPTKTSQKQAPVTRPRRLKINVLIAEKREKER 939
Db 1089 RIVSEKCEVAKGIP-----NVEKTCDEGSSPSKPK-----KGLTLD 1126
Qy 940 KKSVDKDESSSMKISKEGTEIVRPSPKRME-----GDVEKLETPPEKDIAS--STTP 993
Db 1127 PLAEQHDPSSCPAPLKVDEVNAASPQAQNHLEHNGPQGGEDVLTDDMEICTPDRIISP 1186
Qy 994 AKKIKLNRETGKIGNAENASTTKPESEKLESTSKIKOEKYGKAKRRVAGSESS--- 1050
Db 1187 AK-----GEVVSPLAN-----HRLDSPEVNIIPEODECMAHPRAGGEQSSMS 1230
Qy 1051 -----STLVDTSPNMSGGVPYRKSEKTDYKRVYKIMERY-----NNDMTA 1093
Db 1231 SKTLBESGVKQDSSVSVT--SPVETSGKKEGAKQOMLTDWKPLOQVGNISVSTATTS 1288
Qy 1094 PAEDVITIMIOVQSQKWDKDFESEEDVYKTPQIOVGKPSIIKRVTTKPSATAKYTEK 1153
Db 1289 SALDVKALSTVPEVK-----PQGLIEIKSNKV-----RPSGLPDEVKRTAFLNRPRNQ 1339
Qy 1154 ESEQPEKLOKLPKEASHLMQHELRRSSKGSASSEKRAKDRHSGSEKDNPDKRGSAOP 1213
Db 1340 ESSSDQGTTPSRDSDS-----QSRSHSRKRSKSETKSHNRYSYSVSHSRSRSSSTSS 1392
Qy 1214 DKESIVDRLBQGNF-----KTLSSQSK-----ETPRSEKHSYRGSNNDFPGRKQKDY 1265
Db 1393 YRSRYSRSRSDWYSRGTRSRSSSYGSHHRTYSRKR--RSSYDL-HSRRSSTY 1449
Qy 1266 DSRDYSKRRDERELARKKDSPPRGKESLGGQSKLLEEDLPRKGAESKSSSPPR 1325
Db 1450 DSY-YRSRSR--SQRSDSYHRG-----RSYNNRSMGSGSYGSDSS----- 1490
Qy 1326 DKPPHDKAPYETKR 1340
Db 1491 DRSYHHRSPSESSR 1505

RESULT 10
A44265
trithorax homolog HTX, version 2 - human
N:Alternate names: acute lymphoblastic leukemia gene 1 protein (ALL-1)
C:Species: Homo sapiens (man)
C>Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 01-Dec-2000

C:Accession: A44265; A44264; I58112; I37165; I38485
R: Tkachuk, D.C.; Kohler, S.; Cleary, M.L.
Cell 71, 691-700, 1992
A:Title: Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocation
A:Reference number: A44265; MID:93046667
A:Accession: A44265
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-3968 <TKA>
A:Cross-references: GB:I04284; NID:9184393; PIDN:AA58669.1; PID:g184394
R:Gu, Y.; Nakamura, T.; Alder, H.; Prasad, R.; Cnaanani, O.; Cimino, G.; Croce, C.M.; Can
Cell 71, 701-708, 1992
A:Title: The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 g
A:Reference number: A44264; MID:93046668
A:Accession: A44264
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 63-316; GLINSLER, 327, 'Q', 329, 'VR', 332, 'DKETPP', 340, 'T', 342, 'EDKTIVROSPPH
546, 'LQIETSP', 2555-3554, 'N', 3556-3594, 'V', 3596-3899, 'A', 3901-3968 <GU1>
A:Cross-references: GB:I04731; NID:9339921
A:Note: sequence extracted from NCBI backbone (NCBI:117779)
R:Djabelli, M.; Sellieri, L.; Parry, P.; Bower, M.; Young, B.D.; Evans, G.A.
Nature Genet. 2, 113-118, 1992
A:Title: A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute
A:Reference number: I58112; MID:93265134
A:Accession: I58112
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1317-1700, 'DD', 1703-1936, 'H', 1938-2180, 'S', 2182-2328 <DIA>
A:Cross-references: GB:I01986; NID:9307522; PIDN:AA92511.1; PID:9353600
R:Marshall, R.; Grell, J.; Lochner, K.; Nilsson, I.; Slegler, G.; Zweckbrunner, I.; Beg
Br. J. Haematol. 90, 308-320, 1995
A:Title: Molecular analysis of the chromosomal breakpoint and fusion transcripts in the
A:Reference number: I37165; MID:95315013
A:Accession: I37165
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1212-1603, 'GTE' <MAR>
A:Cross-references: EMBL:X83604; NID:9897757; PIDN:CAA58584.1; PID:9899268
R:Gu, Y.; Alder, H.; Nakamura, T.; Schichman, S.A.; Prasad, R.; Cnaanani, O.; Salto, H.;
Cancer Res. 54, 2327-2330, 1994
A:Title: Sequence analysis of the breakpoint cluster region in the ALL-1 gene involved in
A:Reference number: I38485
A:Accession: I38485
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1251-1486, 'G', 1488-1538 <RES>
A:Cross-references: EMBL:U04737; NID:9451554; PID:9451555
C:Genetics:
A:Gene: GDB:ML; HTX; ALL-1; HRX
A:Cross-references: GDB:128819; OMIM:159555
A:Map position: 11q23-11q23
A:Introns: 1338/1; 1362/3; 1406/3; 1444/3; 1493/3; 1525/3; 1566/1
A:Note: the list of introns is incomplete
C:Keywords: acute lymphoblastic leukemia; proto-oncogene; zinc finger
F:1434-1456/Region: zinc finger CCHC motif
F:1479-1506/Region: zinc finger CCHC motif
F:1527-1556/Region: zinc finger CCHC motif
F:1569-1586/Region: zinc finger CCHC motif
F:1873-1900/Region: zinc finger CCHC motif
F:1933-1955/Region: zinc finger CCHC motif

Query Match 4.68; Score 341; DB 2; Length 3968;
Best Local Similarity 20.1%; Pred. No. 1.2e-05;
Matches 283; Conservative 153; Mismatches 463; Indels 512; Gaps 60;

QY 111 LINKFLROAVNNFKNETGTRKRLKRLPPE-----LPLVPPRPLSQNLQ 157
DB 371 LLSNSTRGO-----KKGAOKRIEKEAOLQGRKRYKTVOKNIRPIMPVVAISSRIK 423

QY 158 PRSRSPILROODPVV-----FRYVSPICSDTKTAGSCSD-----SGTSLRLPAP 202
DB 424 TPRFIEDDEYDPDIKARLESTPNRSFS--APSGSSEKSAASQHSQMSSSSSSSP 482
QY 203 SSSLSNSQSS-----LAPVSGNPSSAPAPVDITATVTSVHSEKSDPFPDSOKKL 257
DB 483 SVDTSDQASSEETQVLPERSSTPEVHP-PLP-----ISQSPNESNDRRSRYS---- 532
QY 258 PAAALTEHSHKSGASTAITALMEKEGVCTSPNNSIFVGOSLHGLDLPYTGVRJMAAR 317
DB 533 -----VSERFSGRT----- 542
QY 318 PGGRPEWESHKSLGLVYSPPODIRGERSCYNSINGRHSHRSQRTQSPSLPATPCFV 377
DB 543 -----TKRLSTLQSAPOQ-----QYSSP----- 561
QY 378 PVPPPLPYPPP-----HTLPLPGVPP--PQSPQFPSSQPTTAYS--VPFG 423
DB 562 --PPPLTTPPLPQPPASSISDHT--PWLMPPTIPLASPLPASTABMOGRKRSILEPT 616
QY 424 F-----PPAPANISTACFS-PGV--PTAHSNTMPTTQAPLLSREEFYREQNDKGRES 472
DB 617 FRMTSLKHSRSEPOYSSAKYAKEGILRKPIFNFRPPLTP-----EDVGFAS 665
QY 473 KPPYSGSSYSRSSYTDSSQGLAHIALTSPSAHTLDDLHDPHPPEAEARSAMTVH 532
DB 666 GFASGTAASARLFFSPLHSGTRFDMH--KRSP-----LLRAPRPTPSAHSRIEESYV 716
QY 533 MPRLMDIAHAKRSPPYRRRSRSRSPFERGSPTRKRVPRREKEERYNRRREVPY 592
DB 717 LPSNPTSAGTSSGVSNRKRRKRVFSPINSEPPSP--HSMRRSGHSSLSPLPPPS 774
QY 593 DIKAYVGRSYD-----FRDPFEKERYEMERYKRYKGVVGAQPPR--- 639
DB 775 SVSSLSISVPLATSAIANTPTFFPS-----HSLTQSGEAEKNGPRKQ 819
QY 640 -SANREDF---SPERLLPLIRNSPPTGRREDYAAQSHRNRLGANTPEKIS--TRDSH 694
DB 820 TSAPEEFSSSPPLPFPWFTPGSOTERGRNKKKA-----PEELSKRDAD 865
QY 695 NAKNDPKSKKESENPFGDGKGNKHKHRRRMEKGESESLNP-----ELL 743
DB 866 KSYEKDSRDRDRER---EKERRSRKRRKRGSEIQSSSLALYPRGRYSKRVVGEVY 921
QY 744 ETRKRCGSGGIDETKTDLTFLVPSRDDATPVDEPDMAESIFFKYSVDKDKREKPKV 803
DB 922 ATSSSAKKAATGRKKS-----SHDSGT-----DITSVL----- 950
QY 804 KSDTKRKSDGSAATAKKDNVL-KPSKGPOEKVDGDKESPRSEPLKAKE---EATKID 859
DB 951 -----GDTTAVYTKILIKKRGKNEKTMID--LGP-TAPSLKEKTKLCLSTPSSS 997
QY 860 SVKPSSSS-----QKDEKVTGTRKRAHKSAAKTRROQSPRTFRSRRTYPKTS 907
DB 998 TVHSTSSIGSMLAQADKLPMTPDKRVAASLLKAKKAQDLCTEKRSKQYDQ-----PKAQ 1052
QY 908 SOKSOPRT--RRPRSLRKINYLIAEKNRERKRSKSDVDFESSMKISKVGTETIVP 965
DB 1053 GQSDSDSEFSYVRGR---IKH--VCRRAVALOKRAAVPD----- 1088
QY 966 SPARKMEGVLEKLEPP--EKDKIASSTPPAKIKILNRETGKKIIGNAENASTKEPSEKL 1023
DB 1089 -----DMPTLSALPWEEREKILSS-----MGNDKSSIASGEDEPLAPIKPI 1132
QY 1024 ES-TSSKIKQEK--VGKAKRKYAGSDG-----SSSTLVDTYSTSGSPVAKKSEKTD 1075
DB 1133 KPVTRKKAPODEPPVKKGRARRGCGCPGCOVPDGCVTNCLDKPRFGGKNTIKQDCK-- 1190
QY 1076 TKRTVLTIMEEYNNNDTAPDEVILIMIOVPOSKWDKDDFESEEDVKTQPIQSVKPPSS 1135
DB 1191 -----MRKCONLQWPSKATLOKQAKAVKKKEKKSSTSEKDD-----SKESS 1232
QY 1136 IIKNTVTKPSATKATYKESEQPEKLOLPRKASHLMOHELRRSSKGSASSEKGRADRE 1195


```

Db 421 DQNNASGSDSSSDSEEGSSSSNEDSDQNDY-----DEEDBDVYSEKRRHPEEG 474
QY 79 NSGC-----DECRTTLESDKHCTPTCQNDVSPDALIANKFLQAAYN 122
Db 475 KSSSPGNHREDSENGDKDHEDSSERFSQPTSSH-ETSHSPKDEAVQSRSF--SPLN 531
QY 123 NFKENGY--TRLRKOLPP-----LFLVPPRPLSOR-----154
Db 532 YQSQSGYELESEKEIKQEPSPTTSSASSDLEDMEMPNPLTRMLERHMRPFIDVS 591
QY 155 -----NLQRRS-----PILRQO-----DPVERTVSPQSDT 184
Db 592 FVARIDEIVELNOKARASYEFTGRPPKCNNDVLSIQKIVHEBRDYYYENP--CSEL 650
QY 185 KTA-----GSGSDGTLRLPAPISLITSSOGLAPVSGNSAPARVDITATVSY 240
Db 651 EVAIRDMRKLSDTADLDFRADTDSKELGRDQ---PAGRTSGRPSLDESRTNRLSF-- 703
QY 241 HSEKSDGPRFSDSNKLLPAAALTSEHSKASIAITALEMEKGVPTSPWNSIFVGQSL 300
Db 704 -----DSTHPRALAQSH-----SICIGPM-----TPSTP-----729
QY 301 HGQLPTTGPRINAAR-PEGGRPM-----HSNKLGYLSPPOQIRRGERSYRSIN 353
Db 730 -----PFTSQPLVNTHTLPTSQSTSGITTPRSSQPLMSPVS--RHNSMS-----S 778
QY 354 RGRHNEBROQRQSPSLPATPCFVYRPPRLYPRLPTLPGVYRPPQSPQ--FPSQ 411
Db 779 TGRPAIQTIRHOSVMPRPD--VSIPRPT-PTTIDEMAPRGTPRSKSSSTWVPLNS 834
QY 412 PPRAGSVPPRPPRANISTACFSPVPTASNTMPTTQALSLSEERYRQNDKRE 471
Db 835 PEP-----GPIQLMLTPIVPPHLL-----AAT 859
QY 472 SKFPYSGSYSRSSYDSSGGLAOHIALTLSSAHTLIDLHDHPP--PEAEKRS 528
Db 860 STGTHSVSSAHSSTPRHSISGTPVH-----CEPSNSKT--SQPLPKSRPEKVOIRHD 910
QY 529 MIYHM--PDLMDIAHARSRSPRYRYSRSPPEFR-----GQSP--KRNVP 573
Db 911 TIKSGPSNAIMLQARSQMTSGDKKSNPSTPYVRDAGSDLVQAIMSQPLGLRKL 970
QY 574 REEKEREYENREVPY-----DIKAYGRSVDPRDFEKKERYREWKERYEK 626
Db 971 RIKKSSALONIONHOPRHSNANSTPSTSTHOAMFKKKKKKKKKKKKKKKKKKK 1026
QY 627 YKGVAVGAPRPSANREDSPEPLRLPLNRNSPTFRGREDDVAAQSHNRNLGNYPE 686
Db 1027 -----REARREMKRKETKEER--NRKEMERAKRLED--EEOEKRR-----E 1064
QY 687 KLSTRDSHNAKDPKSKSESENVPGDGKNKHKKRRRNEKGESESFLNEL-LET 745
Db 1065 KKKERE-----RKKEKKVKKAKKEKELKKKKHKK--GSSSDSDSDSDELDDLV 1114
QY 746 SRKCGSSGIDETKTTFLVLRDQATPYRDEPMAESTTFKSVSDKDKREKDKPVKS 805
Db 1115 RKSTKEKT--QEEKDQLALLLKGITIE-----NLKSRRSRDRADHSEKKM 1161
QY 806 DKTKRK-----SDGSATAKKDNVLPKSGQEKYDQREKSPREPLKAKKEBATK 857
Db 1162 QKQOQRVLTLESDDEGGKDGKNGSSNGEESSEKAD--LPPPAPELSAADRKL 1218
QY 858 IDSVKSSSSQDEKVGTPRKAHSAK--DTRQSQPTRRSRK-----TYPKTSSQ 910
Db 1219 V-----LKEKEKELTSSDDEHDNDAGELHQOURLTDRNRKRKOKSLTAYSSDEGERK 1273
QY 911 SQVTRRRRPS-----LRKINTLIAEKNERKKKSVDKP-----947
Db 1274 NVPKRMRDSDSEDAAKHPGWSAKDKQOKRRKLEH--RSESEDSKKNARAKREDFDIP 1329
QY 948 -----ESSMKISKVEGTETLVKPSPKKMGVDKLERTPE-----983

```

```

Db 1330 HEDVSEDETEDEGSRGRSTSTISNTAKE-----RKEKSGKTPRLIPEPTGPL 1382
QY 984 -KDKIASSTTPAKKIKILNETGKIGNAEMASPTKEPSKLESTSS-----KIKO 1032
Db 1383 LSEKILSPKHLSPKSTSTSTKSSISDHENLISPRORNTTSTSTATTSKHEALSTPE 1442
QY 1033 ERYKAKRRKRVAGSESSSTLVDTYSTSGGSPVKRSEKTDKTKYIKT--MEEYNN 1089
Db 1443 KPLSPVYTAKSSVSIDDPSIRDEFMNGSADSPM-----STTGRPVLTAKAMKAFN- 1495
QY 1090 DNTAPADYIIMQVQSMKDDF-----ESEEDYKTPQIQSGKSSIIKNY 1141
Db 1496 -STPRKKEIDQAVQSLFDEEADDEFQYDPFGISTNEKEVSGKDPNIT--KTEPLNNGH 1552
QY 1142 T-----KPSATAKTERES-----EOPKLOKLPKASHE--LMQHELRSSKGSASER 1188
Db 1553 TDLTFSPSSAHASEKQSTKSEDMEDSELYMEKEVMEQVIAQEVAVPPEPSPMEBE 1612
QY 1189 GRAKREHSGSEKNDPKRSGAQDPKESTVDRLSQGHFKTILSQSKET-----RTSE 1242
Db 1613 VKL--ETSPVPEKEPIKMBE--SPROTPTPLDIS-----NNESQTPGAVNNHLHE 1659
QY 1243 KHESVR-----GSSNKDPTGGRKKVYDSDRDS-----SKRRDRGRL 1282
Db 1660 NHDVQTPQLQOPASQHQVAAQSPRPVAPDSQNGPVLVSQOSQSPMSQOSDMAONL 1719
QY 1283 -----ARRKDSPPRGKESLSGQSKLREERDLPRKG-----1313
Db 1720 ILSKINDLAKLHNPEALQATRGDCSGIFQHLLLAQGNQMTPEMLQKAFPA 1779
QY 1314 -----AESKSSNSPPRDKKPHDKAPV-ETKRCEBETKPVDNKSGEREKKA 1360
Db 1780 QOQENENAMQAKMQOQTINKDRIKEQERYVRMYEENRKEVEDREROKKEERORLA 1839
QY 1361 A 1361
Db 1840 A 1840

RESULT 15
T30826
nascnt polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
M:alternate names: alpha-NAC protein
C:species: Mus musculus (house mouse)
C:date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:accession: T30826
R:Yotov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:title: Differential splicing -ln of a proline-rich exon converts alphaNAC into a mus
A:reference number: Z20889; M0ID:96312450
A:accession: T30826
A:status: preliminary; translated from GB/EMBL/DBJ
A:molecule type: DNA
A:residues: 1-2187 <YOT>
A:cross-references: EMBL:U48363; NID:q1666688; PID:q1666689; PIDN:AA018732.1
C:genetics:
A:gene: Naca
A:map position: 10
A:introns: 24/1, 1996/1, 2024/3, 2050/3, 2099/3, 2142/3, 2183/3
A:note: differential splicing converts alphaNAC into a tissue-specific DNA-binding ac
C:keywords: alternative splicing; DNA binding; transcription factor

Query Match 4.4%; Score 327; DB 2; Length 2187;
Best local Similarity 20.9%; Pred. No. 2, 2e-05;
Matches 334; Conservative 182; Mismatches 640; Indels 442; Gaps 77;

QY 8 NMKGAMLTNTGKAITPDAEVAIKKE-----KPPLEPPESSSEDDPTPALLC 61
Db 480 NLPISALVNVG--ADVSPAQAGLPYRKDTTLOPLALAKSESSQS-----ASSLE 529
QY 62 LICKDLMTD-----AVIPLCGNSCDECIKRTTLLESDKHCTCPTCHQNDVSPDALIAN 114

```



```

Db 530 VLSEDTYKKTGGPAPVVPALAGVA-----TTTSLRADSPPAVIRADSCVSPNTV--S 582
Oy 115 KELROAVNNFNKNETGYTKRLRKQLPPLFLVPPRPLSORNOPSRSPIILQOOPVVER 174
Db 583 QPLKRSYVD-----PAMAPRTAKNTAPSTISPLV----- 611
Oy 175 YTVSPFCSDTGTAGSCSDGTLRLPAPSSISLTSNOSGLAPPVSGNPSAPAPYDITA 234
Db 612 -----PLASE-----GCPVASSMAL--SPONASVSETALALSPEI--PKSVFPDPL-A 656
Oy 235 TVSIS-----VHSEKSDGPRSDNKLPPAALITSEHSG-----ASSIATILAMEEK 282
Db 657 EISEFNARKVDVAVSHMESGSSROGH-----PDASVT---AKGTVCCLADSSLDTSVASK 709
Oy 283 GVPGTSPMNSIFVGGSLHGLIPTTGPVRINAARPG-----GRDGEHNSNKLGLVSP 338
Db 710 G-----SALSGASSPLX-PLEVSFLPEAGLAVGPGK--SLNKL--PTPP 750
Oy 339 QOIRGERSCYRSINGRHHSERSORTQSPSLPATPCFVPPV-----PLVPPP 388
Db 751 -----SSKGAVPSTGAPSPKGAPIVPTSSISSKQVPAELILPS 790
Oy 389 PHILP-----LPGVPPPOPSPOF-----PSSQPTAGTSVP----- 420
Db 791 POKTPEEVTAASRLISAVOSPKVDPIMSDVTPTSPPKTSATAVPKDYSATLSLKVPAVYSL 850
Oy 421 -PPGPPAPANISTACFSPGVPATASHN-----TWPTQAPLISKEEYREON 466
Db 851 SPPKAPVAPSNENAT-IVTLEIPTSLKNALAAATPKETLATSLPKVYTS-SPQKTPKSVS 907
Oy 467 DKGRSEKPPYSGSSYSRSSYTDSSQ-----GLAOHI-----HALTLSPSA 506
Db 908 LKGAPAMTSKKAATELAASKDVSPQFPEKEVPLQHVPTSPSPKSVSDTLSCALISPPK 967
Oy 507 AHTLDLHDHPH---PEAEARSAIYAMPDLMDIAHAKRSRPPRYRYSRSPPEFR 563
Db 968 GPATIAETPTYPKPSKPPAKSKTPATPSPE--GVTAVPLEIPPCSKKAPTAAPKESS 1025
Oy 564 GOSPTKRNVPREKEREYFNRRREVP-----PYDIKAYGRSVDFRDPFEKERYREWE 617
Db 1026 ATSSSKR-APKTAVS-----KEIIPSGVTAVPLEISLPLETETSKSATPGESASSP-K 1076
Oy 618 RKYREMYEKYYGVAAGAPRPSANREDFSPERLPLINRNSPFTGRREDYAAGOSHRN 677
Db 1077 RSPK-----TAGPRETPPGGVTAVPPEISLP--PKETPONATPNEISLAASQKRS 1124
Oy 678 RNL-----GG-----NYPEK-----LSTRDSHNAKDNP 701
Db 1125 PKTVPKETPPGCVTAMPLEIPSAPQKAPKTAVPKOIPTPEDAVTILAGSPLSPKKASKT 1184
Oy 702 SKEKSENVPGG---KGNKHKKRRKNEKEGESEFLNPELLETISRKCRGSSGIDE 757
Db 1185 AAPKEAPATPSGVIAVSGEISPSPKTSTKTAAPKENSATLPP---KRSPKTAAPKETPA 1241
Oy 758 TKTDFLFLVPSHDDATPVDEPDMAE---SITFKSVSDKDKREDDKPKVKSDDTKRKS 812
Db 1242 TSSBEGTAVPSISPS-PTPAKSGVPVLTJPKGAPNALAESPASPK-KVPKTAPE 1296
Oy 813 DGSATAKKDNVLKPSKGOEKVGDREKSPRSEPLKAKKEATKIDSVKPSSSSOKDEK 872
Db 1297 ETSTTP-----SPQKIP--KVAGPKFAS--ATPPSKTKPTAV-----PKETSAPSEG 1340
Oy 873 VVG-----TPRKAHSKSAKDTRRQSOPTRRSKRT--VPKTSQSKSQPVTRRRPSLAK 924
Db 1341 VTAVPLEITPPSPRKA-----PKTAPKETPAPEGATTAPVQI--PPSPK 1385
Oy 925 INYLIAEKNEBERKRRKSVDK---DFESSMKISKVEGTE--IVKSPKR----- 969
Db 1386 GSKKAGSKETPTTPEBEGTATAPLEPISSKKTSMASPKETLVTPSSKKLSQTVQPKET 1445
Oy 970 KMEGVEKIELTPEKDKIASSTTPAKKIKLN--RETGKKIGNAENASTTKE-----P 1019
Db 1446 SLEGATAVPLEITPSHKKAPKTVDPKOVLTSPSPKDAPTTL--AESPSSPKKAPKTAAP 1503

```

```

Oy 1020 SEKLEST---SSKIKQEKYKAKARKVAGSEGSSTLVDTYSTSGSPVRK---SEEXT 1074
Db 1504 SERVTTPPEKATPTQKASGTTASKVPPAPAEIQAFAVSSRETPVPVPPVNNPSHKKT 1563
Oy 1075 DTKRTVITMEEYNNNDNTAPAEVDIIMIOVPOSKWDKDFESEEDVKTTOPIQSVGRPS 1134
Db 1564 S-----KTIEL-----KEAPA-----TLPPSPTRKSPKIPSSKRAPRTSAPKEPPASPS 1606
Oy 1135 STIKNVTTKPSATKATYTESEQPEKLOK-----LPKE--ASHELMOHELRS-----SKGS 1183
Db 1607 --IKPVTTSLAQT-----PPSLQKAPSTTIPEKNLAPAVLVSSKSPAAPARAS 1655
Oy 1184 ASSEKGRADREHSGSEKONPDKRKSGA-----OPDKESTVDRLSEOGHFKTLSSQ 1234
Db 1656 ASLSRATAPAPAPKAEATTIPSCKKAATETPIENSTAPSLGAPKETSISVSKVMS 1715
Oy 1235 SKETRTSEKHESVRGS-----SNKDFTPGRDKKYVDYDSRDYSSSKRRDERGEL 1282
Db 1716 PPKKASSSKRASTLPTATLPLSLKEASVLSPTATSSGKDSHISPVSDACSTGTTTPOASER 1775
Oy 1283 ARKDSPPRGKESLS--GOKSKLREERDLPRKGASKSNSSP-----PRDKRPHDKAPY 1336
Db 1776 LPSKGPATFTEMLAAPAESALATTAPLOKSPGANSSASSPPKCPDPBSKR--DTKGLP 1833
Oy 1337 ETKRPEETKPYDKNSGKEREKHAAEARNGKESGANC 1374
Db 1834 SAVALAPQTVPEKDTSK-----AIEFTLLVSPAKGSDC 1866

```

Search completed: September 12, 2002, 19:13:04
Job time: 6549 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 17:49:20 : Search time 25.52 Seconds

(without alignments)
2130.182 Million cell updates/sec

Title: US-09-811-045a-1
Perfect score: 7374
Sequence: 1 MMEVXPNNKGMALTRTGKY.....WRRSWLLGRRRAPSSRNPS 1404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 353 | 4.8 | 3969 | 1 HRX_HUMAN | Q03164 homo sapien |
| 2 | 342 | 4.6 | 1453 | 1 NRCR_MOUSE | P30415 mus musculu |
| 3 | 337.5 | 4.6 | 1020 | 1 NEH_HUMAN | P12036 homo sapien |
| 4 | 333.5 | 4.5 | 2468 | 1 MAPB_HUMAN | P46821 homo sapien |
| 5 | 326 | 4.4 | 2492 | 1 ATRX_HUMAN | P46100 homo sapien |
| 6 | 321 | 4.4 | 1391 | 1 MST2_DROXY | Q08696 drosophila |
| 7 | 321 | 4.4 | 2464 | 1 MAPB_MOUSE | P14873 mus musculu |
| 8 | 317 | 4.3 | 1462 | 1 NRGR_HUMAN | P30414 homo sapien |
| 9 | 315 | 4.3 | 633 | 1 MLH_TETTH | P40631 tetrahymena |
| 10 | 315 | 4.3 | 3866 | 1 HRX_MOUSE | P55200 mus musculu |
| 11 | 305 | 4.1 | 1822 | 1 ZAP3_HUMAN | P47550 homo sapien |
| 12 | 302 | 4.1 | 2842 | 1 APC_RAT | P70478 rattus norv |
| 13 | 299.5 | 4.1 | 2459 | 1 MAPB_RAT | P15205 rattus norv |
| 14 | 299 | 4.1 | 2476 | 1 ATRX_MOUSE | Q61687 mus musculu |
| 15 | 294 | 4.0 | 699 | 1 NP14_HUMAN | Q91760 homo sapien |
| 16 | 291 | 3.9 | 1359 | 1 ATRX_CAERL | O91760 homo sapien |
| 17 | 290 | 3.9 | 2805 | 1 MAPA_HUMAN | P78559 homo sapien |
| 18 | 288.5 | 3.9 | 2845 | 1 APC_MOUSE | O61315 mus musculu |
| 19 | 286.5 | 3.9 | 1790 | 1 USOL_YEAST | P33386 saccharomyc |
| 20 | 285 | 3.9 | 705 | 1 TRDN_RABIT | Q28820 oryctolagus |
| 21 | 283.5 | 3.8 | 488 | 1 CYL2_BOVIN | Q28092 bos taurus |
| 22 | 282.5 | 3.8 | 831 | 1 NPH_RAT | P16884 rattus norv |
| 23 | 279 | 3.8 | 728 | 1 TRDN_HUMAN | Q10061 homo sapien |
| 24 | 278.5 | 3.8 | 1189 | 1 YOH6_YEAST | P47035 saccharomyc |
| 25 | 277.5 | 3.8 | 667 | 1 CYL1_BOVIN | P33662 bos taurus |
| 26 | 276.5 | 3.7 | 1210 | 1 AF4_HUMAN | P51825 homo sapien |
| 27 | 276.5 | 3.7 | 2142 | 1 BAT2_HUMAN | P46634 homo sapien |
| 28 | 276 | 3.7 | 771 | 1 CALD_CHICK | P19257 gallus gall |
| 29 | 275.5 | 3.7 | 872 | 1 S3B2_HUMAN | Q13435 homo sapien |
| 30 | 274 | 3.7 | 2843 | 1 APC_HUMAN | P25054 homo sapien |
| 31 | 273.5 | 3.7 | 704 | 1 NP14_RAT | P41777 rattus norv |
| 32 | 273.5 | 3.7 | 1781 | 1 ARAQ_HUMAN | Q02952 homo sapien |
| 33 | 273 | 3.7 | 2774 | 1 MAPA_RAT | P34926 rattus norv |

ALIGNMENTS

| RESULT | ID | HRX_HUMAN | STANDARD: | PRT: | 3969 AA |
|--------|-------------|---|-----------|------|---------|
| AC | Q03164 | Q14845; Q16364; Q13743; Q13744; Q90WA3; | | | |
| DT | 01-OCT-1993 | (Rel. 27, Created) | | | |
| DT | 01-NOV-1993 | (Rel. 32, Last sequence update) | | | |
| DT | 01-MAR-2002 | (Rel. 41, Last annotation update) | | | |
| DE | | Zinc finger protein HRX (ALL-1) (Tritheorax-like protein). | | | |
| GN | | MLL OR HRX OR ALL1 OR TRX1 OR HTRX. | | | |
| OS | | Homo sapiens (Human). | | | |
| OC | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | | Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. | | | |
| OX | | NCBI_TaxID=9606; | | | |
| RN | | [1] | | | |
| RP | | SEQUENCE FROM N.A. | | | |
| RX | | MEDLINE=9304667; PubMed=1423624; | | | |
| RA | | Tkachuk D.C., Kohler S., Cleary M.L.; | | | |
| RT | | "Involvement of a homolog of Drosophila trithorax by 11q23 | | | |
| RT | | chromosomal translocations in acute leukaemias."; | | | |
| RL | | Cell 71:691-700(1992). | | | |
| RN | | [2] | | | |
| RP | | SEQUENCE FROM N.A. | | | |
| RX | | MEDLINE=9629053; PubMed=8703835; | | | |
| RA | | Nilson I., Loechner K., Stiegler G., Grell J., Beck J.D., Fey G.H., | | | |
| RA | | Marschalek R.; | | | |
| RT | | "Exon/intron structure of the human ALL-1 (MLL) gene involved in | | | |
| RT | | translocations to chromosomal region 11q23 and acute leukaemias."; | | | |
| RL | | Br. J. Haematol. 93:966-972(1996). | | | |
| RN | | [3] | | | |
| RP | | SEQUENCE OF 1-1909 FROM N.A. | | | |
| RX | | MEDLINE=93390935; PubMed=8378076; | | | |
| RA | | Yanamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S., | | | |
| RA | | Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.; | | | |
| RT | | "Two distinct portions of LFG19/ENL at 19p13 are involved in t(11;19) | | | |
| RT | | leukemia."; | | | |
| RL | | Oncogene 8:2617-2625(1993). | | | |
| RN | | [4] | | | |
| RP | | SEQUENCE OF 1317-2328 FROM N.A. | | | |
| RC | | TISSUE=Brain; | | | |
| RX | | MEDLINE=93265134; PubMed=1303259; | | | |
| RA | | Djabali M., Seilieri L., Parry P., Bower M., Young B.D., Evans G.A.; | | | |
| RT | | "A trithorax-like gene is interrupted by chromosome 11q23 | | | |
| RT | | translocations in acute leukaemias."; | | | |
| RL | | Nat. Genet. 2:113-118(1992). | | | |
| RN | | [5] | | | |
| RP | | SEQUENCE OF 1251-1538 FROM N.A. | | | |
| RX | | MEDLINE=94215165; PubMed=8162575; | | | |
| RA | | Gu Y., Alder H., Nakamura T., Schlichtman S.A., Prasad R., Canaan O., | | | |
| RA | | Saito H., Croce C.M., Canaan E.; | | | |
| RT | | "Sequence analysis of the breakpoint cluster region in the ALL-1 gene | | | |
| RT | | involved in acute leukemia."; | | | |
| RL | | Cancer Res. 54:2326-2330(1994). | | | |
| RN | | [6] | | | |
| RP | | SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B). | | | |
| RX | | MEDLINE=95322025; PubMed=7598802; | | | |
| RA | | Mbangkollo D., Burnett R., McCabe N., Thirman M., Gill H., Yu H., | | | |

| | | | | | |
|----|-------|-----|------|--------------|---------------------|
| 34 | 272 | 3.7 | 1220 | 1 IFP2_HUMAN | O60841 homo sapien |
| 35 | 271.5 | 3.7 | 1386 | 1 ZAP3_MOUSE | O91017 mus musculu |
| 36 | 271.5 | 3.7 | 2116 | 1 MYS2_DICDI | P08799 dictyosteli |
| 37 | 270.5 | 3.7 | 793 | 1 CALD_HUMAN | O05682 homo sapien |
| 38 | 270.5 | 3.7 | 1233 | 1 YF16_YEAST | P43597 saccharomyc |
| 39 | 269.5 | 3.7 | 598 | 1 CYL1_HUMAN | P33663 homo sapien |
| 40 | 269 | 3.6 | 1253 | 1 DSPP_HUMAN | O9nzv4 homo sapien |
| 41 | 266 | 3.6 | 1087 | 1 NPH_MOUSE | P19246 mus musculu |
| 42 | 264.5 | 3.6 | 1374 | 1 RNC_HUMAN | O9nrr4 homo sapien |
| 43 | 263.5 | 3.6 | 1130 | 1 YL17_CAERL | O11102 caenorhabdi |
| 44 | 262 | 3.6 | 1898 | 1 TRHY_HUMAN | O07283 homo sapien |
| 45 | 261.5 | 3.5 | 1581 | 1 PPRB_HUMAN | O15648 h petroxisom |

RA Rowley J.D., Diaz M.O.;
 RT "The human MLL gene: nucleotide sequence, homology to the Drosophila
 RT trz zinc-finger domain, and alternative splicing";
 RL DNA Cell Biol. 14:475-483(1995).
 RN [17]
 RP SEQUENCE OF 1212-1603 FROM N.A.
 RX MEDLINE=95315013; PubMed=7794749;
 RA Marschalek R., Grell J., Lochner K., Nilsson I., Slegler G.,
 RA Zwechbrunner I., Beck J.D., Fey G.H.;
 RT "Molecular analysis of the chromosomal breakpoint and fusion
 RT transcripts in the acute lymphoblastic SEM cell line with chromosomal
 RT translocation t(4;11)",
 RL Br. J. Haematol. 90:308-320(1995).
 RN [8]
 RP SEQUENCE OF 1421-1540 FROM N.A.
 RX MEDLINE=94020842; PubMed=8414518;
 RA Forster A., Rabbitts T.H.;
 RT "A method for identifying genes within yeast artificial chromosomes:
 RT application to isolation of MLL fusion cDNAs from acute leukaemia
 RT translocations.";
 RL Oncogene 8:3157-3160(1993).
 RN [9]
 RP CHROMOSOMAL TRANSLOCATION WITH GAS7.
 RX MEDLINE=20183971; PubMed=10706619;
 RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
 RA Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
 RA Williams T.M., Lange B.J., Felix C.A.;
 RT "Detection of leukemia-associated MLL-GAS7 translocation early during
 RT chemotherapy with DNA topoisomerase II inhibitors,"
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
 RN [10]
 RP CHROMOSOMAL TRANSLOCATION WITH AF3P21.
 RX MEDLINE=20115194; PubMed=10648423;
 RA Sano K., Hayakawa A., Piao J.-H., Kosaka Y., Nakamura H.;
 RT "Novel SH3 protein encoded by the AF3p21 gene is fused to the mixed
 RT lineage leukemia protein in a therapy-related leukemia with
 RT t(3;11)(p21;q23).";
 RL Blood 95:1066-1068(2000).
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 CC T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLT1/ENL;
 CC T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLT2/AF4; T(9;11)(P22;Q23)
 CC THAT INVOLVES MLL AND MLT3/AF9; T(6;11)(Q27;Q23) THAT INVOLVES
 CC MLL AND MLT4/AF6; T(11;17)(Q23;Q21) THAT INVOLVES MLL AND
 CC MLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLT7/AFX1;
 CC T(10;11)(P12;Q23) THAT INVOLVES MLL AND MLT10/AF10;
 CC T(1;11)(Q21;Q23) THAT INVOLVES MLL AND AF1Q; T(11;19)(Q23;P13.3)
 CC THAT INVOLVES MLL AND ELN; T(11;19)(Q23;P23) THAT INVOLVES MLL
 CC AND GAS7; and t(3;11)(p21;q23) that involves MLL and AF3P21.
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
 CC -1- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL; L04284; AAA58669.1; -;
 DR EMBL; Z69744; CAA93625.1; -;
 DR EMBL; Z69745; CAA93625.1; JOINED.
 DR EMBL; Z69746; CAA93625.1; JOINED.
 DR EMBL; Z69747; CAA93625.1; JOINED.
 DR EMBL; Z69748; CAA93625.1; JOINED.
 DR EMBL; Z69749; CAA93625.1; JOINED.
 DR EMBL; Z69750; CAA93625.1; JOINED.

DR EMBL; Z69751; CAA93625.1; JOINED.
 DR EMBL; Z69752; CAA93625.1; JOINED.
 DR EMBL; Z69753; CAA93625.1; JOINED.
 DR EMBL; Z69754; CAA93625.1; JOINED.
 DR EMBL; Z69755; CAA93625.1; JOINED.
 DR EMBL; Z69756; CAA93625.1; JOINED.
 DR EMBL; Z69757; CAA93625.1; JOINED.
 DR EMBL; Z69758; CAA93625.1; JOINED.
 DR EMBL; Z69759; CAA93625.1; JOINED.
 DR EMBL; Z69760; CAA93625.1; JOINED.
 DR EMBL; Z69761; CAA93625.1; JOINED.
 DR EMBL; Z69762; CAA93625.1; JOINED.
 DR EMBL; Z69763; CAA93625.1; JOINED.
 DR EMBL; Z69764; CAA93625.1; JOINED.
 DR EMBL; Z69765; CAA93625.1; JOINED.
 DR EMBL; Z69766; CAA93625.1; JOINED.
 DR EMBL; Z69767; CAA93625.1; JOINED.
 DR EMBL; Z69768; CAA93625.1; JOINED.
 DR EMBL; Z69769; CAA93625.1; JOINED.
 DR EMBL; Z69770; CAA93625.1; JOINED.
 DR EMBL; Z69771; CAA93625.1; JOINED.
 DR EMBL; Z69772; CAA93625.1; JOINED.
 DR EMBL; Z69773; CAA93625.1; JOINED.
 DR EMBL; Z69774; CAA93625.1; JOINED.
 DR EMBL; Z69775; CAA93625.1; JOINED.
 DR EMBL; Z69776; CAA93625.1; JOINED.
 DR EMBL; Z69777; CAA93625.1; JOINED.
 DR EMBL; Z69778; CAA93625.1; JOINED.
 DR EMBL; Z69779; CAA93625.1; JOINED.
 DR EMBL; Z69780; CAA93625.1; JOINED.
 DR EMBL; D14540; BAA03407.1; -;
 DR EMBL; L01986; AAA92511.1; -;
 DR EMBL; U04737; AAA18644.1; -;
 DR EMBL; S78570; AAB34770.1; -;
 DR EMBL; X83604; CAA58584.1; -;
 DR EMBL; S66432; AAB28545.1; -;
 DR EMBL; AF231988; AAG26332.2; ALT_TERM.
 DR TRANSFAC; T02337; -;
 DR MIM; 159555; -;
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR003889; Fyrich.C.
 DR InterPro; IPR003888; Fyrich.N.
 DR InterPro; IPR001965; PHD.
 DR InterPro; IPR003616; PostSET.
 DR InterPro; IPR001214; SET.
 DR InterPro; IPR002857; Znfx-CXXC.
 DR Pfam; PF00856; PHD; 3.
 DR Pfam; PF02008; Zfx-CXXC; 1.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00542; FYRICH; 1.
 DR SMART; SM00541; FYRICH; 1.
 DR SMART; SM00249; PHD; 4.
 DR SMART; SM00508; PostSET; 1.
 DR SMART; SM00317; SET; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50280; SET; 1.
 KW Proto-oncogene: Chromosomal translocation; DNA-binding; Bromodomain;
 KW Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;
 KW Alternative splicing.
 KW
 FT DOMAIN 17 102
 FT DNA_BIND 169 180
 FT DNA_BIND 217 227
 FT DNA_BIND 301 309
 FT ZN_FING 1147 1194
 FT ZN_FING 1431 1482
 FT ZN_FING 1484 1533
 FT ZN_FING 1566 1627
 FT DOMAIN 1703 1748
 FT DOMAIN 3840 3969
 FT DOMAIN 137 143
 FT DOMAIN 561 564
 FT DOMAIN 568 571
 FT SITE 1362 1363
 A.T HOOK (BY SIMILARITY).
 A.T HOOK (BY SIMILARITY).
 A.T HOOK (BY SIMILARITY).
 CXXC-TYPE.
 PHD-TYPE 1.
 PHD-TYPE 2.
 PHD-TYPE 3.
 BROMODOMAIN (DIVERGENT).
 SET.
 POLY-GLY.
 POLY-PRO.
 POLY-PRO.
 BREAKPOINT FOR TRANSLOCATION TO FORM MLL-

Query Match 4.8%; Score 353; DB 1; Length 3969;
 Best Local Similarity 20.2%; Pred. No. 2.6e-06;
 Matches 289; Conservative 155; Mismatches 466; Indels 522; Gaps 61;

```

QY 90 LLESDKHTCPTCHQNDVSPDALIANKFLROAVNNFKNETGYKRLRKOLPPF----- 141
DB 360 IIPSSKRT-----DATIAKOLLORA-----KKGAQKTEKEAQAQGRKVKQ 402
QY 142 -----LFLVPPRPPLSQNRLOPRSRSPILROODPVY-----FRYTSPTCSDKT 186
DB 403 VKNIKQFIMPVVAISRIKTPRFRIEDEDYDPKIAKRLRESTPNRSRFS-APSCGSSSEK 461
QY 187 AGSCSD-----SGLSRLLPAPISISLTSSNOS-----LAPVSGNPSAPAPVDITATV 236
DB 462 SSAASQHSQMSDSSSRSSSVDTSTDSQASEEIQVLPEERSDTPPEVHP-PLP-----I 515
QY 237 SISVHSEKSDGPFPRDSDKLLPAAALTEBHSKGASSIAITALEMEKGVPGTSPMNSIFVG 296
DB 516 SOSPENESNDRSRYS-----VSBRPFSRPT----- 542
QY 297 QSLHGLIPTTGPVIRIARPGGGRPGWENSKLGYLVSPPOQIRKGRSCYSINRGR 356
DB 543 -----TKKLSTLQSAPOQ----- 555
QY 357 HHSERSORTOSPLPATPCFVVPVPPPLYPPP-----HTLLPLPGVPP--PQNSP 405
DB 556 -----QTSSSP-----PPPLLPPLLPLOPASSISDHT--PWLMPPPTPLASP 595
QY 406 QEPSSQPTAGYS--VPPPGF-----PPAPANISTAFCS-PGV--PTAHSTMTPT 451
DB 596 FLPASTAMOGKRKSLIREPFRMTSLKHSREPOYFSSAKYAKGLIRKPIFDNPRPP 655
QY 452 QAPLISREFFREONDKRESKFPYSSSYSSSTTDSQGLAQHITHALTLSPSAHTLD 511
DB 656 LTP-----BDVCFASGFSASGTAASARLPSPLHSGRFTPMH--KRSP----- 695
QY 512 LLDHDPHPPEAEASAMIVHMDLMDIAHARSRPYRYSRSRSPREFGOSPTKEN 571
DB 696 LLRAIRFTFPEAHSHKIFESVTLPSKRTSAGTSSSGVSNKRKKRYFSPIRSEPRSPS--H 753
QY 572 VPREEKEREYFNRYREVPDYDIKAYGRSVD-----FRDPFEKERYREWEKRYR 621
DB 754 SMKTRSGRLSSSELPLTPPSSVSSLISVPLATSALEPFTPPS----- 800
QY 622 EWEKYTYGAYGAQPRP-----SANREDF--SPERLLPLNRNSPFTTGREDTAAGOS 674
DB 801 --HSLTQGESAEKNQRPKQTSAPAEPPSSSSPTPLPFWPTPGSQTERGRNKDKA---- 854
QY 675 HNRNMGNYPEKLS-TRDSHNAKNPKSKERESENVPGDGKGNHKKRKRRENEKGE 733
DB 855 -----PELSKDRDADKSVKEDKSRERDRER-----EKENKRESKKEKRRKKGSEIQ 900
QY 734 SESFLNP-----ELLETSRCKRGSSGIDETKTDTLTVLPSRDPATPVREDPMDA 782
DB 901 SSSALYPGRVSKYKVEGVDATSSAKKATGRKKS-----SHDSGT-----DI 945
QY 763 ESITKYSVSDRKREKDKPKVSKTKRKSDSATAKKUNVL-KFSKGQOEVDGDREMS 841
DB 946 TSVTL-----GDTTAVKTKLLIKGKGLKGTMTND--LG 977
QY 842 PSEPEPLKAKE---EATKIDSVKPPSSS-----OKDEVOTSPKRAHSKAK 886
DB 978 P-TASLSLEKTKLCISTPSSSTVKHSTSSIGSMLAODLPLMTDKRYASALLKAKAQLCK 1036
QY 887 DTRROSQPTRSRKRTVPTSSQKQPVPT--RPRSLRKIVYLLAREKNEBEKRRKSV 944
DB 1037 IKSSKSLKQTDQ-----PRAQOQESDSETSVGRPR--IKH-VCRRAAVALLGRRAVE 1086
QY 945 KPESSSMKISIKVEGEIYKPSPKRKMEDVDEKLETP--EKDKLASSTTPAKKILANE 1002
DB 1087 PD-----DMPTLSALPWEEREKILIS-----MGND 1111

```

```

QY 1003 TGKRTGNAENASTKPESEKLES-TSSKIKQEK--VKGAKRKRVAQSDG-----SSSTLV 1054
DB 1112 DKSSIASGEDEAPLAPLPIKPIKPVIRNKARQPEPPYKGRKGRSCGCGCCQVPEDCGCT 1171
QY 1055 DYTSTSTGSGSVPRKSEKTDKRTVYIKTMEYNNNDTAPADVTIMIQVPOSKWDKDF 1114
DB 1172 NCLDPRKFGGRNIRKQCK-----MRCKQMLQWMPKAVYLQKAKAVKREKRSK 1221
QY 1115 ESEEDVKTQPIQSVGRPSSIIKNVTTKPSATFAYTEKESQPEKLOKLPKASHELMQ 1174
DB 1222 TSEKRD-----SKESSVAKNVV-----DSSQKp----- 1244
QY 1175 HFLRSKGSASSEKGRADREHSGSEKNDPKRKSQAOPDKSTYVRLSEOGHFTLSOS 1234
DB 1245 -----TPSAREDPARKSSSEPPPKPPEKSEBEENVAPGE 1282
QY 1235 SKETRTSEKHESVKGSSNKF-----TPGRDKKVDYDSRDYSSSKRRDERGELARRKD 1287
DB 1283 SKQATTPASRKSSKQVSOQALVIPPQPTTGPPRK-----EVPKTPP 1324
QY 1288 SPPRGKESISGKSKLREERDLPKKGAEKSKNSSPPRDKKHDHKAPEYETRRPCEETKP 1347
DB 1325 SEPKRKQP-----PPESGPEQSKQKVAAPRPSIPEVOK-PKEKERP-----PP 1367
QY 1348 VDK-----NSGKER-----EKHAEAARNGKESG 1371
DB 1368 VKKQENAGTLNLTLSNGNSSKQKIPADGVIRIRVDFEEDCEAEVWEMG 1419

```

```

RESULT 2
NCRR_MOUSE NCRR_MOUSE STANDARD; PRT; 1453 AA.
AC P30415;
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE NK-tumor recognition protein (Natural-killer cells cyclophilin-
DE related protein) (NK-TR protein).
GN NKTR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93133624; PubMed=8421688;
RA Anderson S.K.; Gallinger S.; Roder J.; Frey J.; Young H.A.;
RA Ortaldo J.R.;
RT "A cyclophilin-related protein involved in the function of natural
RT killer cells ";
RL proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC STRAIN=BALE/C; TISSUE=Blood;
RA Anderson S.K.;
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
CC INVOLVED IN THE FUNCTION OF NK CELLS.
CC - SIMILARITY: CONTAINS 1 CYCLOPHILIN-LIKE PLIASE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: 104289; AAA37500.2; ALT_INIT.
DR HSP: Q27450; IA33.
DR MGD: MGI:97346; NKTR.
DR InterPro: IPR002130; CSA_PLIase.
DR Pfam: PF00160; pro_isomerase; 1.

```

[illegible]

| | RESULT | 3 |
|------------|--|-------------------------|
| NEFH_HUMAN | | |
| ID | NEFH_HUMAN | STANDARD; PRT; 1020 AA. |
| AC | P12036; | |
| DT | 01-OCT-1989 (Rel. 12, Created) | |
| DT | 01-OCT-1989 (Rel. 12, Last sequence update) | |
| DT | 16-OCT-2001 (Rel. 40, Last annotation update) | |
| DE | Neurofilament triplet H protein (200 kDa neurofilament protein) | |
| DE | (Neurofilament heavy polypeptide) (NF-H). | |
| GN | NEFH OR NEFH. | |
| OS | Homo sapiens (Human). | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | |
| OX | NCBI_TaxID=9606; | |
| RN | [1] | |
| RP | SEQUENCE FROM N.A. | |
| RP | MEDLINE=88328981; PubMed=3138108; | |
| RA | Lees J.F., Shneidman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.; | |
| RT | "The structure and organization of the human heavy neurofilament | |
| RT | subunit (NF-H) and the gene encoding it."; | |
| RL | EMBO J. 7:1947-1955(1988). | |
| CC | -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M | |

CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PWM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PWM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X15306; CAA33366.1; JOINED.
 CC EMBL: X15307; CAA33366.1; JOINED.
 CC EMBL: X15308; CAA33366.1; JOINED.
 CC EMBL: X15309; CAA33366.1; JOINED.
 CC PIR: S00979; QFH0H.
 CC MIM: 162230; QFH0H.
 CC InterPro: IPR001664; IF.
 CC Pfam: PF00038; filament; 1.
 CC PROSITE: PS00226; IF; 1.
 CC Intermediate filament; Coiled coil; Neurone; Phosphorylation.
 CC FT DOMAIN 1 100 HEAD.
 CC FT DOMAIN 101 413 ROD.
 CC FT DOMAIN 414 1020 TAIL.
 CC FT DOMAIN 101 132 COIL 1A.
 CC FT DOMAIN 133 145 LINKER 1.
 CC FT DOMAIN 146 244 COIL 1B.
 CC FT DOMAIN 245 266 LINKER 12.
 CC FT DOMAIN 267 288 COIL 2A.
 CC FT DOMAIN 289 292 LINKER 2.
 CC FT DOMAIN 293 413 COIL 2B.
 CC SQ SEQUENCE 1020 AA; 111780 MW; 1177C9DCB3DCFD1D4 CRC64;

Query Match 4.6%; Score 337.5; DB 1; Length 1020;
 Best Local Similarity 21.4%; Pred. No. 2.7e-06;
 Matches 212; Conservative 164; Mismatches 458; Indels 155; Gaps 43;

QY 443 AHSNMPPTQAPLLSREERYRQNDKGRSKFPYSSSYSSSYSDSSQGLAQHIALT- 501
 DB 120 AHNRSLEGSAAL-----RQQAQGRSA---MGLYERE--VREMGAVALRLGAANG 165
 QY 502 -----LSPSAATLIDLHDHPHPPEEAARSAMIVMP-----DLMDIAHARS 545
 DB 166 QRLBEGHLLLEIDIAHYRQLDDEARQREAEAAARLAFQAQEAARVDLQKKAALQE 225
 QY 546 SPYRRYRSRSRSPPEFGC-----SPTKRVNPREEKEREYFNRYREVPPEYDIKAVY 598
 DB 226 ECGYLIR-RHNOEVEGLLQIQSGAAQOMQAEIRDAIKCVTSALREIRAQLE----- 279
 QY 599 GRSVDFRDFEKKERYREMYREKRYKGAIVAGQRPSPNRRDPSPERLLPLNIRN 658
 DB 280 GAHV--QSTLQSE--EW--FVRLDRLSEAKAVNTDMARSQOEITEYRQLQARTTE 331
 QY 659 SPFTGRREDYVAGOSH---RNRLGANTPEKLTSDS--HNAKDPKSKSESENVPGD 713
 DB 332 LEAKTKTSLERQSELEDHRQADIASYQDAIQDIAELRNTKWMMAQLREYQDL---- 388
 QY 714 GGNKHKHKKRRRNEKGESEFLNPELLETLSRKCR-GSSGIDETKKTJTLFVLPGRDA 772
 DB 389 -----LNVKALDIEIAVYAKILLE-GECCRIIGGPIPSLPELPIPSVSTH 435

QY 773 TPVRDEPMDAESTITFKSVSDK-----KREKDKPKYKSDKTKRKSQDSATKAKDNVLKPSK 828
 DB 436 IYKXSE---EKIKYVESKEFIVVEQTEETQYTEETEE---EKPAKEEGKEEG 488
 QY 829 GPOEKDGDRE--KSPRSE--PPLKA---KEEA-TKIDSKSSSSQKDEKYTGPR 878
 DB 489 GDEEKEGDEETKSPAPAEASPEKEAKSPYKEAKSPAEPKEAKSPAEPKSP 548
 QY 879 KAHKSAKTRQOSQPRTRRSKRYVPKTSQSKSQPYRTTRPSLKRINTLIREKNERK 938
 DB 549 KAKSPKPEAK--SPEAKSPKEAKSPAEPKSPKAKSP--AKKSPKPAKSP 603
 QY 939 RKKSVDKDES-----SSMKIKVDETEIVK-----SPRKMGDEVEKLETPPEKDKIS 989
 DB 604 AKSPYKEAKSPAEPKSPKPAEAKSPAEPKSPKPAEAKSPKPAEAKSPK 662
 QY 990 ---STTPAKKIKLNRGTGKIGNAENASTTKPESEKLESTSSKIKQKYGAKRRVAGS 1046
 DB 663 PEKAKSPYKAEAKSPKAKSPYKAEAKSPKAKSPYKEAKS---PEKAKSPYKEAKSP 719
 QY 1047 EGSSTLVDTYSTSTSGSPVR---KSEKTDTKRIVITMEYKNNDNTPADVTIMQ 1103
 DB 720 ERAKSPYKEEAKTPPEKAKSPYKAEAKSPKAKSPKPA-RTLVKSPKATPAKE--EAR 775
 QY 1104 VPQSKDKDQDFSEEDVKTTO---PIQVGPSSIIKNVTTKPSATAKYTEKESEQPE 1159
 DB 776 SPADKTPPEKAKSPYKEEVSPKAKSPKADAKAP--KEIRKEEVKSPVKEEKPQEV 833
 QY 1160 KLQKLPEASHLMOHELRLSSKGSASSEKGRAKDRHSGSEKDNP--DKRRSGAOPDKE 1216
 DB 834 KYKEPKKAEKEE-----KAPATPKTEKKDSKEEAPKAEKAPKVEKEKPEAVEKPE 887
 QY 1217 STYDRLSEQGHKRTLSQSSKEKRTSKHESVSGSSKQDTPGDDKVVYDSDRYSSSKRR 1276
 DB 888 SKVEAKKEE-----AEDKKVPTPEKEAPKAVEVEDAKPKKETEVAKKEPDADAKAK-- 939
 QY 1277 DERGELARRKDSPPRKESLSQSKLREERDLPKGAESKSSNSPPRD-KKPHDKAP 1335
 DB 940 -EPSKAKEKKEAPPEKDI-KEEKAKEKPEK--PKTEAKAKEDDKLSNEPSKPAKMAE 995
 QY 1336 YETKRPCETTKPVYKNSGKEREKHAPEAR 1364
 DB 996 KSSSTQKDSKPEKAT---EDKAKAKG 1020

RESULT 4
 MAPB_HUMAN
 ID MAPB_HUMAN STANDARD; PRT; 2468 AA.
 AC P46821;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain LC1].
 GN MAP1B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retal brain;
 RX MEDLINE=95104835; PubMed=7806212;
 RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
 RT "Cloning of human microtubule-associated protein 1B and the
 RT identification of a related gene on chromosome 15.";
 RL Genomics 22:273-280(1994).
 CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
 CC STABILIZING MICROTUBULES.

```
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -1- PWM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (by similarity).
CC -1- SIMILARITY: TO MAP1A.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: I06237; AAA18904.1; -.
CC
CC DR MIM: 157129; -.
CC DR InterPro: IPR000102; MAP1B_neuraxin.
CC DR Pfam: PF00414; MAP1B_neuraxin. 10.
CC DR PROSITE: PS00230; MAP1B_NEURAXIN; 6.
CC KW Microtubules; Repeat; Phosphorylation.
CC FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.
CC FT REPEAT 1878 1894 MAP1B 1.
CC FT REPEAT 1895 1911 MAP1B 2.
CC FT REPEAT 1912 1928 MAP1B 3.
CC FT REPEAT 1929 1945 MAP1B 4.
CC FT REPEAT 1946 1962 MAP1B 5.
CC FT REPEAT 1963 1979 MAP1B 6.
CC FT REPEAT 1997 2013 MAP1B 7.
CC FT REPEAT 2014 2030 MAP1B 8.
CC FT REPEAT 2031 2047 MAP1B 9.
CC FT REPEAT 2048 2064 MAP1B 10.
CC FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
CC KKEE AND KKEI/V REPEATS).
CC
CC SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;
```

```
Query Match 4.58; Score 333.5; DB 1; Length 2468;
Best Local Similarity 22.18; Pred. No. 9e-06;
Matches 170; Conservative 127; Mismatches 282; Indels 191; Gaps 33;

QY 678 RNLEGNP-----EKLSTRDSHNAKDNP--KSKESKES-ENVPDGGKGGKHKHR 723
DB 519 KDLTGQPTPVYKQTKLKQRADSRSLKPAKRLPSKSVKRESKEERPEYTKVNHVERPP 578
QY 724 KRNEEKGESESLNPELLETSTRKCGSGIDETKTD-TLEVLPSRDDATPVADPEMDA 782
DB 579 KVESKEK-----VMVKKDKPVK-----TETKPSVTEKEVPSKEEPSVKAEEAEK 623
QY 783 ESTFKEVSODKREKDKPKYKSDTKRKSDGSAATKADNVLKSKGQEVDDREKSP 842
DB 624 QADIVKPKAAAEKTKVKEKTPKDEKKEK--KPKKEVAKKDKTPPIK-----EKKP 675
QY 843 RSEPLKAKEEATKIDSVKPSSSODEKVTGTPRAKSHSADTRRQSPRTRRSKRT 902
DB 676 KKEEVKKEVKEKKEKKEKKEKKEKKEKKEK-----TPPEVKEVKEV-----KEEKE 720
QY 903 VPKTSSQKSPVTRRRPSLKIYLIAREKNEKREKRSKSDVKDFESSMKISVEGETEI 962
DB 721 VKK-----EKEPEKKEIKLIPKAKKSSPPLSEAKKPPAA 754
QY 963 VKPEPKRMGGDVKELEKRTPEKOKIASSTPAKIKILNREGKIKGNAENASTHKEPSEK 1022
DB 755 LKPK-----VPKKEESVKKDSVAAG-----KKEPK 779
QY 1023 LESTSKIKOEKVGAKRKVAGSESSSTLVDTSTSGSGPVKRSSEKTDTRKRVYK 1082
DB 780 -----GKIYIKKKGKAAEAENAAVGTGATL---AAVMAAGIAIGPAKLELEKRSIMS 831
```

```
QY 1083 TMEYENND-NTAPAEVYIMIQV-POSKWDKDFESEEDVKTTOPIOS--VGKPSSTIK 1138
DB 832 SPEDLTDFEELKAEKEDVDTMDIKPQLELIED-----EKKKEIPEPAVYIQKREYTK 886
QY 1139 NVTTKPSATYAKYTEKES--QPEKLQKLPKREASHLMQHELRSSKSSASSEK----- 1189
DB 887 GPASPEGLITTTGEGEGCEQTPPELPEVEKQGVADIEKFEDEAGGESESETDYEEKA 946
QY 1190 ---RAKREHSGSE-----KDNP-----DKRKSQAQNP--KESIVDR 1221
DB 947 ETEBAEPEDEDEEHVCVSAKSHSPTEDEESAKAEADATIREKRESVAASGDRAEEDPDE 1006
QY 1222 LSEQGHFKTLQSQSKETRTSEKHSYVGSNNKDPPTGRRDKKVYD---DSRDYSSSKRRD 1277
DB 1007 AIEKGE---AEQSEADEDEDKADAR---EEVEPEKMEAEDEVMAVVKAAAGAGAE 1060
QY 1278 ERGLARRKDSPPR--GKESLSGOKSKLREERDLPKKAESKKSNSPP--RDKRPDHKA 1334
DB 1061 QYGFLT-----TPRKQLAQSFGREPASSIHDETLP--GGSESEATASDEENEDQPEEFTA 1115
QY 1335 ---PYETKRRCPEETKPVDKNSGKEREKHAAEARNKGSSGANCAYLTLR 1380
DB 1116 TSGYGTSTIEISSEPTPWDEMS--TPROVMSDETNNETESPQSFEVNIWK 1164

RESULT 5
ATR_X_HUMAN STANDARD; PRT; 2492 AA.
ID ATRX_HUMAN
AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-HX).
CN ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND VARIANT S-1860,
RP AND VARIANTS ATR-X.
RX MEDLINE=97123494; PubMed=8966741;
RA Pickelits D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrelli O.W.J.,
RA Gibbons R.J.;
RT "ATR-X encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RN Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE=97386582; PubMed=9244431;
RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
RA Collaen L., Schwartz C., Fontes M.;
RT "determination of the genomic structure of the XNP/ATRX gene encoding
RT a potential zinc finger helicase.";
RN Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE=95179111; PubMed=7874112;
RA Strayton C.L., Dabovic B., Gulisano M., Gecz J., Broccoli V.,
RA Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RN Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050;
RA Gecz J., Pollard H., Consalez G., Villard L., Strayton C.L.,
RA Millaesau P., Khrestchatsky M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PKGI in Xq13.3.";
RN Hum. Mol. Genet. 3:39-44(1994).
```


[5]
 RN SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
 RP MEDLINE=95211835; PubMed=7697714;
 RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
 RT "Mutations in a putative global transcriptional regulator cause X-
 RL linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
 RN Cell 80:837-845(1995).
 [6]
 RP SEQUENCE OF 1375-2492 FROM N.A.
 RA Pearce A., Chapman J.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 [7]
 RP E2H2 BINDING.
 RX MEDLINE=98167853; PubMed=9499421;
 RA Cardoso C., Timmit S., Villard L., Khrestchatskiy M., Fontes M.,
 RL Colleaux L.;
 RT "Specific interaction between the XNP/ATR-X gene product and the SET
 RL domain of the human E2H2 protein.";
 RN Hum. Mol. Genet. 7:679-684(1998).
 [8]
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RP HETEROCHROMATIN.
 RX MEDLINE=20040663; PubMed=10570185;
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
 RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
 RT "Localization of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RL chromosomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
 [9]
 RP DISEASE.
 RX MEDLINE=20213147; PubMed=10751095;
 RA Villard L., Fontes M., Ades L.C., Geetz J.;
 RT "Identification of a mutation in the XNP/ATR-X gene in a family
 RT reported as Smith-Fineman-Myers syndrome.";
 RL Am. J. Med. Genet. 91:83-85(2000).
 [10]
 RN VARIANT ATR-X SER-1713.
 RX MEDLINE=97196774; PubMed=9043863;
 RA Villard L., Lacombe D., Fontes M.;
 RT "A point mutation in the XNP gene, associated with an ATR-X phenotype
 RL without alpha-thalassemia.";
 RL Eur. J. Hum. Genet. 4:316-320(1996).
 [11]
 RN VARIANT JM GLN-2131.
 RP MEDLINE=96224392; PubMed=8630485;
 RA Villard L., Geetz J., Mattel J.-F., Fontes M., Saugier-Verber P.,
 RA Munnich A., Lyonnet S.;
 RT "XNP mutation in a large family with Joubert-Marsidi syndrome.";
 RL Nat. Genet. 12:359-360(1996).
 [12]
 RN VARIANTS ATR-X.
 RP MEDLINE=97467722; PubMed=9326931;
 RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asembauer B.,
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
 RA Levin M.L., Masuro M., Nerl G., Plempont M.E., Slaney S.F.,
 RT "Mutations in transcriptional regulator ATRX establish the functional
 RT significance of a PHD-like domain.";
 RL Nat. Genet. 17:146-148(1997).
 [13]
 RN VARIANT ATR-X LEU-246.
 RP MEDLINE=20123062; PubMed=10660327;
 RA Fichera M., Romano C., Castiglia L., Falla P., Roberto C., Amata S.,
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;
 RT "New mutations in XNP/ATR-X gene: a further contribution to
 RT genotype/phenotype relationship in ATR/X syndrome.";
 RL Hum. Mutat. 12:214-214(1998).
 [14]
 RN VARIANT SHS LYS-1742.
 RP MEDLINE=99347960; PubMed=10417298;
 RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,

RA Prieto F., Fontes M., Martinez F.;
 RT "Mutation of the XNP/ATR-X gene in a family with severe mental
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:
 RT demonstration that the mutation is involved in the inactivation
 RT bias.";
 RL Am. J. Hum. Genet. 65:558-562(1999).
 [15]
 RP VARIANT CMS THR-2050.
 RX MEDLINE=99326061; PubMed=10398237;
 RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
 RA Curtis M.;
 RT "Carpenter-Waziri syndrome results from a mutation in XNP.";
 RL Am. J. Med. Genet. 85:249-251(1999).
 [16]
 RN VARIANTS ATR-X E-175; 178-V-K-198 DEL; S-190; P-219; L-246 AND C-249.
 RX MEDLINE=99219535; PubMed=10204841;
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belongue J.,
 RA Lossi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,
 RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;
 RT "Evaluation of a mutation screening strategy for sporadic cases of
 RT ATR-X syndrome.";
 RL J. Med. Genet. 36:183-186(1999).
 [17]
 RN VARIANTS ATR-X S-179; L-190; I-194; C-246; F-1552; S-1645 AND C-1847.
 RP MEDLINE=20451413; PubMed=10995512;
 RX Wada T., Kubota T., Fukushima Y., Saitoh S.;
 RA "Molecular genetic study of Japanese patients with X-linked
 RT alpha-thalassemia/mental retardation syndrome (ATR-X).";
 RL Am. J. Med. Genet. 94:242-248(2000).
 CC -I- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
 CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
 CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
 CC -I- SUBUNIT: PROBABLY BINDS E2H2. BINDS ANNEXIN V IN A CALCIUM AND
 CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY
 CC similarity).
 CC -I- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
 CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
 CC INTERACTING WITH HP1.
 CC -I- ALTERNATIVE PRODUCTS: 5 ISOFORMS: 1, 2, 3, 4 (SHOWN HERE) AND 5;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -I- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -I- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
 CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
 CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPLETING SEVERE
 CC PSYCHOMOTOR RETARDATION, FACIAL DYSMORPHISM, UROGENITAL
 CC ABNORMALITIES, AND ALPHA-THALASSEMIA. AN ESSENTIAL PHENOTYPIC
 CC TRAIT ARE HEMOGLOBIN H ERYTHROCYTE INCLUSIONS.
 CC -I- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF CARPENTER-WAZIRI
 CC SYNDROME (CMS), AN X-LINKED RECESSIVE CONDITION CHARACTERIZED BY
 CC MODERATE MENTAL RETARDATION, SHORT STATURE, BRACHYDACTYL WITH
 CC EXCESSIVE SKIN CREASES, AND WIDENING OF THE KNUCKLES.
 CC -I- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JOUBERT-MARSIDI SYNDROME
 CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY
 CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,
 CC MICROGENITALISM AND EARLY DEATH.
 CC -I- DISEASE: DEFECTS IN ATRX ARE A CAUSE OF SMITH-FINEMAN-MYERS
 CC SYNDROME (SFM). CLINICAL FEATURES INCLUDE SEVERE MENTAL
 CC RETARDATION, MICROCEPHALY, GROWTH FAILURE, FACIAL ANOMALIES AND
 CC BILATERAL CRYPTORCHIDISM. DUE TO THE CLINICAL OVERLAP WITH ATR-X
 CC SYNDROME, SOME PATIENTS ORIGINALLY DIAGNOSED AS HAVING SFM, MIGHT
 CC BE AFFECTED BY A VARIANT OF ATR-X SYNDROME WHICH LACK HEMOGLOBIN H
 CC INCLUSIONS.
 CC -I- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF SUTHERLAND-HAAN X-LINKED
 CC MENTAL RETARDATION SYNDROME (SHS). IT IS CHARACTERIZED BY SEVERE
 CC MENTAL RETARDATION WITH SPASTIC PARAPLEGIA, MICROCEPHALY, SHORT
 CC STATURE AND CRYPTORCHIDISM.
 CC -I- SIMILARITY: BELONGS TO THE SNEF/RAD54 HELICASE FAMILY.
 CC -I- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).

CC -----
 DR EMBL: U72937; AAB49970.2; -
 DR EMBL: U72938; AAB49971.2; -
 DR EMBL: U72935; AAB40698.1; -
 DR EMBL: U72904; AAB40698.1; JOINED.

Query Match 4.4%; Score 326; DB 1; Length 2492;
 Best Local Similarity 19.0%; Pred. NO. 1.7e-05;
 Matches 287; Conservative 227; Mismatches 566; Indels 434; Gaps 63;

62 LICKDITDVIIPCCGSSCDECI-----RTTLESDEKHTCPTCHQNDVSPDALI 112
 229 LICODF-----CHNACCKCICLNLNLRKELSTIMENDNOMWCICH-----PEPL 274
 113 ANKFLQAVNN-EKNEGYTKRLKQLPPLFLVPPRPISQRLQPRSPILLROODPV 171
 275 D---LVYACNVEEN-----LEOLLQKKKIKIVDSEKSNK 307
 172 VFRYT-----VSPYCS--DTKTAGSCSDGTLR---LP-----APSSISLTJNS 212
 308 VYEHSTSFSPKRTSSNCGEKKLDSCSGSVITYSYALLVPEKIMAKKALLETIANM 367
 213 S-----LAPPVSGNPSSAPAPVDITATVSISSHSEK-----DGPR--DSDNKL 257
 368 SSYVFKLQKQNDSEISSATKRLQKAFKSVLADIKKAHLALEDLNEFRADAVNK-- 425
 258 PAALITSEHKGASSTIYTLAMEKGVPTSP-----WNSITFGOSLILHQ 303
 426 -EKNTKEHK-----VIDAKFETKARKGKPCALEKKDISSEAKLSRKQVSEHMH-Q 476
 304 LIPYTGVRINARPGGGR-----PGWHSNKLGYL---VSPPOI----- 341
 477 NVLTE-EQRNKSTGSGHKHSDRKEEPQYRPANTSELDMDIYVSPSVEDITFENLETA 535
 342 -----RGRERS-----CYRSINRGRHNSQR--TQSPSLPATPCFV 377
 536 MEYOSSVDHQDGSSTGEQEVESVYKLTISKNRGIKSKTKAKYTKELYKLP--V 593
 378 PVPPRPILYPPPHLPLP-----GYPP----- 401
 594 SLPSNPTKAGDCQEVDPQDKYKSGLNPKLEKGLQENSDEHLENVSLLESDEL 653
 402 QFSPQF---PSSOPPTAGYVPPPGFPAPANISTACFSGVPTAHNTMPTQAPILSR 458
 654 RBSRPVKTTLRKRTETN-----PVTNSNDEEC-NETVAREKOKLSVP-----R 696
 459 EEEFYREONDKG---RESKPPYSGSSYSRSSYTDSSQGLAQHIALTLSPSAHTLIDLH 514
 697 KKDKNSSDSALINDPKRNKLPKSKQSETVDQNSDEMLALILNGVSMSSSSSDDTINE 756
 515 DHHHPPEAARASAMVHMDLMDIAHARSRPYRKRYSRSPPEF---RGOSPTKRN 571
 757 IH-----TNHKTLYDLKTQAKDKGKKRRKSSSTSGSFDYTKKRSKSSSI 802
 572 VPBEKEKREYFNRYREVPPYDIKAYYGRSVDPDPPEKRRYRMEKRYEWKEYKGY 631
 803 ISKKKROT-----SESNYDSELEKE----- 824
 632 AVGAOPRANREDSPERLLPLINRNSPTGRREDYAGAGSHRNRLNGNPEKLSTR 691
 825 -IKSMKIGIAR---TJKRILP-----NTK 845
 692 DSNANAKDNPKSKKESEBNVGDGKGNKHKHRRNRNEKGESESLINPELLETSRKRG 751
 846 DFSSSEDEKHSK-----KGMDOGHKMLTKSQEESDDAEKROK--RETFSSAEG 893
 752 SSIDETKDTLFLVPSRDATPVRD--EPMDAESTFKVSDKDKRE-KDKKRVASDKT 808
 894 TVDKITTIMELRDLPLPKQOASASTDGVDLGSKSEOSTSLEVRKVAETREKSKHLTKT 953

QY 809 KRKSDGSATAKDNVILKPSKQPEKYVDGDKRSPRSEPLTKAKAEATKIDSKVP---SSS 866
 954 CKRVQGLSDIARKPLKKQDSDTSESD-DKKQSKKGEEKKKSDFPKKVTIKMEQYESS 1012
 QY 867 SQKDEKVTGTPRAH--SKSKADTRROSQPTRRSKRVPTSSQKSOQVPTRRPRLRK 925
 1013 SDGTEKLPEREELICHPEKPGKQIKQIKNGTTDDEKSKKIRDTSSKKDE-----L 1060
 QY 926 NYLIAREKNERKRRKSVDPDPESSSMKISKVGTETIVKSPRRK--MEGDVLEKLETP 983
 1061 SDYAEKSTGDCDSDSEDKSKNGATGKREKCKLIGKSSRRKQDSSSDTKY---SMK 1118
 QY 984 KDIASSTTPAKRIKL-----NRETGKIGNAENASTKEPSE---KLESTSSKIKQ 1032
 1119 EDGCSNDRKRLRIELERERNLSSKRRTKEIQGSSSDAEFESSDKKKKQRTSSKKA 1178
 QY 1033 EKYKGR-----AKRVAGSESSSTLVDTYSTSGSPRKSPEKIDTYRYIKTM 1084
 1179 VIVKEKRNLSRTSTRKQDITSSSSSDIEDDQNSIGGSSDEOKIKPTENLVLSH 1238
 QY 1085 EEEYNNDTAPAEVIMIQYQSKWKD-----DESEEDYKTTQPIQ 1128
 1239 TGFCQSSGDALSKSVYTVDDDDDDNDPENRIAKMLLEIKANLSDSDGSSDEPEE 1298
 QY 1129 SVGKPSIINKVTTKOSATAK-YTEKESQPEKLOKLPKASHLMOHELRSKGSASSE 1187
 1299 --GKKRTGKQNEENPGDEEAKNOVNSESDSDSESKPR-YRRRLIRHKLTVDGSGEE 1355
 QY 1188 KGRKAKREHSGSEKDNDRKSGAQPDKESTVRLSROGHFKTLSSQSK---TTRSEK 1243
 1356 K-RTKPEH--KEVKGGRNRKVSSEDESD---FQSGVSESVSESDQRPRTSAKK 1409
 QY 1244 HESVRCSSNKDFTPGBDKKYDYSRDYSSSKRRDERGELARRDSPPRGKESLQSKSL 1303
 1410 AE-----LEENRGY---KOKKKRRIRIVQDSSSENNKSEEBEEEK 1449
 QY 1304 REERDLPKKAESK---SNSSPPRDKP---HDKAPYTKRPCEETKRVVKNSGK 1354
 1450 EEEEEEEEEEEEDENDSDSKSPGKRRKIRKILKDDKLTETQNALKEE-----E 1501
 QY 1355 EREKHAARNGKE 1368
 DB 1502 ERRKRIARERERE 1515

RESULT 6
 MST2_DROHY
 ID MST2_DROHY STANDARD: PRT: 1391 AA.
 AC 008696;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Axoneme-associated protein mst101(2).
 GN MST101(2).
 OS Drosophila hydei (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7224;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=95045538; PubMed=7957199;
 RA Neesen J., Padmanabhan S., Buenemann H.;
 RT "Randomly arranged repeats of a novel, highly charged 16-amino-acid
 RT motif representing the major component of the sperm-tail-specific
 RT axoneme-associated protein family Dmst101 form extended
 RT alpha-helical rods within the extremely elongated spermatozoa of
 RT Drosophila hydei";
 RL Eur. J. Biochem. 225:1089-1095(1994).
 CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
 CC SPERMATIDS.


```

Db 525 SOSYSGSSRSRTASSSSHSRGSRSSSKSGHRRKASPRKRTASOLSEKKPYKTEP 584
QY 384 L-----YPPPHPLLP--GVPPQ-----ISPPQSSQ-----411
Db 585 LRATMAQNNVAVVAVANINIPVPLSDSPPPSRMRPGOKPMRPSYERIOEMAKTTHTLL 644
QY 412 PPTAGSVPPGPPAPPAINISTACFSPVPLANSNMPPTQALRLSREERYRQNDKGRE 471
Db 645 PLOSTYSL-----ANIK-----ETGSSSYHKREKSESDQSTYSYSDRSSE 667
QY 472 SKFPYSGSSYSRS-----SYTDSQGLAQHIALHTLPSA-AHTLDLHDHPH-----518
Db 688 SS-PRSRSSRSRSRSRYT-RSRSLASS-HRSRSRSPSRSSSRKKYSDHSCSRSSSY 744
QY 519 ---PPEPARSAMIVHMDLMDIAH-----ARSRRPYRRYRSRSRS 558
Db 745 TSLSDGRRARRKRLSSGKKNKSHKSSSEKTLHKKYVGRGRSSGCVRFY-SESR 803
QY 559 PPEFRG---QSPTRKNVPPREKEKEFENRYREVPPYDIKAYYGRGVDPDEPREKRYRE 615
Db 804 SLDYSSESSQVATQSAQEKQ-----GQMERTHNKOENKNGEE 845
QY 616 WERYKREMYEKYKGYAVGAQPRPSANREDFSPERLLPLN---INSPFTGRREDYAA 671
Db 846 KSKSEKREC-----PHSKKRLTKENLSDHLKNG-SKPKRKNYAG 882
QY 672 ---GQSHRNRLGNYPEKLTROSHNAKADNPKSEKSENVPGDGKGNKHKRRRN 777
Db 883 SKWDSSESNBERDVTKN-----SKNDSH---PSSDKEDEGAT-----915
QY 728 EEKGESEFLNPELLETSTRKRCGSGIDETKDTLFLVPSRDA---PPVRDEPMDAES 784
Db 916 ----SSES-----EVSIEHAKVPT-----TKSNTSLPDDNGAKSSQKRTSTDSSE 962
QY 785 IFFKSVSDKDKREK---DKPKVSDKTKRRKSDGATAKKDNVLPKSGPOEYVDKREK 840
Db 963 SCNSNSNNNGKPKQKHGSKENLKRHTTKVKE-KLKGRKD---KHKKAPK-----RKQ 1012
QY 841 SPRSEPLKAKKEAKIKDSVKRSSSSQKDEKVTGTPRAHKSAD---TRQSQPRT 896
Db 1013 AFHMOPPLEGEEEBEEDID-KOVTOSEKKEKVS---ENNETIDNLIKTKSSSEDL 1066
QY 897 RRSKRTVPKTSQSQPVPTRRPRSLKINLYLAREKNEREKRRKGVDDKFESSMKIK 956
Db 1067 SGKHDIVTVSSD-----LDQFTKDDSKLSIPALMTTEENVAC 1104
QY 957 VEGTEIVKQSPKRMGQVEKLETERPEKDIASTTPAKKIKLRNRTGKKIGNA-----1010
Db 1105 LQNIQHYEES---VPNGEDVLQTDNMEIC---TPDRSSPAKVETSPLGNARLDTPD 1157
QY 1011 ----ENASTTKEPSEKLESTSKIOEKYKGAKKRVAGSESSSTLVDTYSTSGGS 1065
Db 1158 INIVLQDMATEHPQAEVVKQESSMSSEKVLGE---VGKODSSS---ASLASGES 1207
QY 1066 PVKSESEKIDTKRTVITKMEYN-----NDNTAPAEVYIIMIOVPOSKWPKDFESE 1117
Db 1208 TGKKEVAEKSQILIDKKKPLQGVGNLAPNAAITSSAIVEKVLTVYPEAK-----PQGL 1262
QY 1118 EEDVKTQPIQSYGKPSSTIKVNTTPSATAKTERESEQPEKLOKLPKASHLMQHEL 1177
Db 1263 RLRIKSKNNV---RPGSLFDEVRKTAFLNRBRNQESSDEQTPSRDDSS-----QS 1311
QY 1178 RSKSGASSEKGAKEHNSGSEKNDPKRKSQAQPDKESTVRLSEQGHF---KTLISQ 1234
Db 1312 RSPSRSRSSSRKSRRTSVSTSHSRSRSSSTSSYSRSRSRSGWYSGKRTSRSS 1371
QY 1235 S-----KETRTSEKHSVNGSSNKKDFTTP-GRDKKVDYDSSSKRRDRGELARRKDS 1288
Db 1372 SSVRSYKSHHTSSRSRS-RSSS---YDPRHSRSYTYDST-YRSRSRSR-----SQRSDS 1422
QY 1289 PPGKASLSGQSKLREERDLPKKGAABSKSSNSPPDKKPHDKAPYETTKR 1340
Db 1423 YHRG-----RSYNNRRSRSCRSYGSDES-----DRSYSHHRSPESSSR 1460

```

```

RESULT 9
ID MLH_TETTH STANDARD; PRT; 633 AA.
AC P40631;
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Micronuclear linker histone polypeptide (MIC LH) [contains:
DE Micronuclear linker histone-alpha; Micronuclear linker histone-beta;
DE Micronuclear linker histone-gamma]
GN MLH.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymena; Tetrahymena.
OX NCBI_Taxid=5911;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-CU401;
RX MEDLINE=94088506; PubMed=8264578;
RA Wu M., Allis C.D., Sweet M.T., Cook R.G., Thatcher T.H.,
RA Gorovsky M.A.;
RT "Four distinct and unusual linker proteins in a mitotically dividing
RT nucleus are derived from a 71-kilodalton polypeptide, lack p34cdc2
RT sites, and contain protein kinase A sites."
RL Mol. Cell. Biol. 14:10-20(1994).
CC -1 SUBCELLULAR LOCATION: NUCLEAR; MICRONUCLEI.
CC -1 PTM: ALL FOUR HISTONES ARE PROCESSED FROM THE PRECURSOR MOLECULE.
CC THEY ARE PHOSPHORYLATED IN GROWING AND DIVIDING CELLS BUT NOT IN
CC NONROWING (STARVED) CELLS. THE N-TERMINAL OF ALPHA AND DELTA IS
CC BLOCKED.
CC -1 SIMILARITY: CONTAINS 2 HMG BOXES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M87306; AAC18874.1; -.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.
DR SMART; SM00398; HMG; 2.
KW Nuclear protein; Chromosomal protein; DNA-binding; Repeat;
KW Phosphorylation.
FT CHAIN 1 399 MICRONUCLEAR LINKER HISTONE-ALPHA.
FT FT 1 199 MICRONUCLEAR LINKER HISTONE-DELTA.
FT CHAIN 200 399 MICRONUCLEAR LINKER HISTONE-GAMMA.
FT CHAIN 400 633 MICRONUCLEAR LINKER HISTONE-BETA.
FT DNA_BIND 12 74 HMG BOX 1.
FT FT DNA_BIND 96 164 HMG BOX 2.
FT DOMAIN 181 391 LYS/SER-RICH.
SQ SEQUENCE 633 AA; 70624 MW; F3B44B/AAE7B3P0C CRC64;

Query Match 4.3%; Score 315; DB 1; Length 633;
Best local Similarity 21.6%; Pred. No. 1,2e-05;
Matches 167; Conservative 126; Mismatches 283; Indels 196; Gaps 28;

QY 650 RLPLPLNR--NSPFTGRREDYAAQSHRNRLGNYPEKLTARD---SHNAK-----697
Db 4 RRTKLNIQPRKPPS---NTYQAFVLEKKNALGKKNNDKKVQADYDNKLNNNEKRLQKL 59
QY 698 -DNPKSKEK-----SENVPGDGKGNKHKRRKRNDEKEESESFLNPELLETSTRKRCG 751
Db 60 VDAAEKRYKELEFHYNNHIOGKR-KQYVQVAVYPEKPKPISSSF---RFLDENQKTA 115
QY 752 SSGIDETKDTLFLVPSRDAFPVRDEPMDAES-----ITFKSVSDKDKREKDKPK 802
Db 116 AKKHDLTNAKILKIMSEDFNNLPOKEVKYVEDAYQYEAQYLVEFKMKNKYGQAQAKQ 175

```

```

OY 803 VSDTKTKRSKSDGSAKKNVLPKSGPOEKVDGDEKSPRPPLKAKAEATKIDSX 862
DB 176 TRKNSTSKSRSSSSKSGSV---SKG-----RTKSTSK---RADSSASQGSOS 221
OY 863 PSS-----SSODEXT-GTPRKANS-----KSAKOTRRSQPPTRRSKRTVPTSS----- 908
DB 222 SSSNRRKSSSSDQKGTSSSRKASNSKGRKNTSKRSSSSSSSSSSSSSSSKN 281
OY 909 OKSQPRTRRPSRLKINLILAREKNERERKKSVDKDESSSMKISKVEGTIVKPSPK 968
DB 282 KSSSSSKGKSSSR-----GKKASSSKNRKSKSKDRSSSSSKGRK----- 323
OY 969 RMEGVEKLETPPEKOKLASSSTTPAKKIKLNFETGKKIGNAENASTTKPESKLESTSS 1028
DB 324 -----SSSSSKNRKASSSRGKSSSSSGKRSKSSK---QERKNSHAD 363
OY 1029 KIKQEVKKGAKKAVAGSESSSTLVDTYSTSGSPVKKSEKEDTKTVAKTMEYN 1088
DB 364 TSKQMEDGOKRR-----OSSSAKDESSSKSR-----NSMKKEARTKA----- 404
OY 1089 NDMTAPAEVDYIMIOVQSKWDKDFEEDVDYKTTQPIQSVKPPSSIIKNVTKPSATA 1148
DB 405 NKMSA-----SKASKSGSKSKSAS-----KSKGSSSKGKNKSSRSASAKP 446
OY 1149 KYTEKESEDEPEKLQKLPKASHLMOHILRSSKGSASSEKGRADREHSGSEKDNPDKR 1208
DB 447 KSNAAQ-----SNNTHQTADESSENASTQTRTGRGROEKQDMVNEKSNKSSS 496
OY 1209 SGQOPDKESTVDLSEGHKTKTSSOSK-----EPTSEKHSVAGSSKDKPTPG 1258
DB 497 KGRKNSKSNTRSKSKSKSASKSRKSKKDDTTNNGRFRSKSRSSKSKSEAPNPSN 556
OY 1259 RDKKVDSDYSSSKRDRGELARRKDDPPRGKESLQSKSLREERDLPKGAESK 1318
DB 557 KMEYIEPKREESDRKRRRESOSAK-----TSQKSKNR-----SDSK 597
OY 1319 SNSPPDKRPHDKADYETKRCPEETKPYDKNSGKREKHAAPARGKSS 1370
DB 598 MTAEDPKKNNNAEDSK-----GKKRKEGKTGAGKAKAN 630

```

```

CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
CC -1- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: L17069; AA62593.1; -.
CC MGD: MG1:96995; M1.
CC InterPro: IPR001487; Bromodomain.
CC InterPro: IPR003889; Ftyich_C.
CC InterPro: IPR003888; Ftyich_N.
CC InterPro: IPR001965; PHD.
CC InterPro: IPR003616; PostSET.
CC InterPro: IPR000093; Recr.
CC InterPro: IPR001214; SET.
CC InterPro: IPR002857; Znf-CXXC.
CC Pfam: PF00628; PHD; 3.
CC Pfam: PF00856; SET; 1.
CC Pfam: PF02008; zf-CXXC; 1.
CC SMART: SM00297; BROMO; 1.
CC SMART: SM00542; FYR; 1.
CC SMART: SM00541; FYRN; 1.
CC SMART: SM00249; PHD; 4.
CC SMART: SM00508; PostSET; 1.
CC SMART: SM00317; SET; 1.
CC PROSITE: PS50014; BROMODOMAIN_2; 1.
CC PROSITE: PS50280; SET; 1.
CC DNA-binding; Nuclear protein; Zinc-finger; Metal-binding;
CC transcription regulation; Alternative splicing; Polymorphism.
KW NON_TER
FT DNA_BIND 67 78 A.T HOOK (BY SIMILARITY).
FT DNA_BIND 115 125 A.T HOOK (BY SIMILARITY).
FT DNA_BIND 199 207 A.T HOOK (BY SIMILARITY).
FT ZN_FING 1044 1091 CXXC-TYPE.
FT ZN_FING 1330 1381 PHD-TYPE 1.
FT ZN_FING 1383 1432 PHD-TYPE 2.
FT ZN_FING 1465 1529 PHD-TYPE 3.
FT DOMAIN 1605 1650 BROMODOMAIN (DIVERGENT).
FT DOMAIN 3737 3846 SET.
FT DOMAIN 35 41 PRO-RICH.
FT DOMAIN 459 469 PRO-RICH.
FT DOMAIN 1231 1238 POLY-PRO.
FT DOMAIN 3533 3536 POLY-GLU.
FT DOMAIN 3693 3697 POLY-GLU.
FT VARSPIC 1503 1505 MISSING (IN ISOFORM 2).
FT VARIANT 1497 1497 K -> T.
SQ SEQUENCE 3866 AA; 420976 MW; ADFCS5E14E806FID CXC64;

Query Match 4.38; Score 315; DB 1; Length 3866;
Best Local Similarity 19.88; Pred. No. 6.8e-05;
Matches 257; Conservative 147; Mismatches 421; Indels 472; Gaps 54;

```

```

Db 425 RSRRYS-----MSERSEGSRA----- 440
QY 308 TGPRIINARGGGRPGMEHSHNKIGLYVSPQQLRGERGRCYRSINCRHHSEKQRTQS 367
Db 441 -----TRKLPLOSAPO-----OKTSS 457
QY 368 PSLPATECFVPPPPPLYPPLPPHTLPLPPGVPPQSPQSPSSQPPNAGYVPPPGFPYA 427
Db 458 S-----PPPLPLTPPP-----PL-----QPASGIDHPWMLMP- 485
QY 428 PANISTACSFSGVPTANSNTPTQAPLLSREEFYREONDGKRESKPEYSSGYSRS--S 485
Db 486 -----PTIPLA-SPFLPASAAPMOGRKSL-----REPTFRWTSLKHSRSPQ 528
QY 486 YTDSSQGLAQIHAHLTSPSAHLLDLHHPPEE-----AASAMIVH 532
Db 529 YFSSAKYAKE-----GLIRKP-----IFDNFRPPLPEDVGFASGFSASCTASARLFSPLH 581
QY 533 MPDLMDIAHARS-----RSPPYRRYRSR-----SPPEFGQSPTKRNVPREKEKREYFN 583
Db 582 SGTFRDI-HKRSPLLRAPRTFPAHSHKIFESVTLPSKRTSSGASGSSGVNKRKKRYFS 640
QY 584 RYREVP--PPYDIKAYYGR-----SVDFRDPF 608
Db 641 PIREPSPSPSHMRTSRGRLSTSELPLTPPSSVSSLSIPVPLASALNPTFTFPHS 700
QY 609 EKEVYREMERKYREMYEKYYGYAVGAQPPRANREDFSP--RLPLNINSPTTRRR 666
Db 701 LTQSGDSTEKQR-----ARKQTSAPAEPPSSNSPALFPWFTPGSQTEKGRK 747
QY 667 EDYAGQSHRRNRLNGVNYPEKLS--TRDSHNAKDNPKEKSEVNPQGDGKHKHKKHRK 725
Db 748 KDTA-----PELSKDRADKVEKDKSERRER-----EKNKRESREK 789
QY 726 RNEKGESESEFLNP-----ELLETSKRCGSSGIDETKTDTLFLVPSRDATP 774
Db 790 RKKSQSDIQQSSALPYGVKSEKAVAGEVGTSSAKAKATGRKSSS-----LDSGADVAP 844
QY 775 VRDPMMAESTTFKSVSDKDKREKDKPKVKDKTKRKSQDATAKKNVL-KPSKGPQEK 833
Db 845 V-----TLGDTTAVKAKILLKKGKGNLEK 868
QY 834 VDGQ-----REKSP-RSEPLKKAKEEATKIDVSPSSS--QKQEDVTGPRKHA 881
Db 869 NNLDLGAAPBLEKERPTCLASPSSTYKHSSTISGMLAADLPMTDKRYASILKAK 928
QY 882 SKSAKDRROSQPTRRSKRTVPPTSQKSQPVPT--RRPR--SLRKINYLAREK--- 933
Db 929 AOLCKIEKSKLKQTDQ-----PRAQGEQSDSEFVSVPRIKHVYRAAVALGKRAVF 983
QY 934 -----NREKRRKSVDKDFESSMKISKVGEETLYKP-SP-----KR 969
Db 984 PDDMPTLSALPWEREREKILSSMGDDKSS-----VAGSEDAEPLAPIKIPYTRKA 1037
QY 970 KMEGDVEKLEKT-----PEKDKIASSTP----- 993
Db 1038 POEPPVAKGRKSRRCGCGPCQVPEDCGICTNCLDKPRFGGRNLIKQCKKMKCONLQMW 1097
QY 994 AKKIKLNETGKRIKGNAMENASTKEPEKLESTSKIKOEKVGKA-----KRVAGS 1046
Db 1098 PSKASLQKOT-KAVKKKKKKSKTTEKKESKSTSVKSPLEPAQKAAPPREAPKPKSS 1156
QY 1047 EGSSSTLVDTYSTSGSPV-----RKSEKTDTKRTVIKTMERYNDN 1091
Db 1157 EPPRRKPEVE--EKSEEGGAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 1210
QY 1092 TAPAEVIMIQVOSKMDKDFESEEDVYTPQIOSVGKPSSTIKINVTIKPSTATKYT 1151
Db 1211 TAPQKK-----EAPKA-----VPSEPKKQPPPEPGEQSKQKVAADLPSTPVKQK 1257
QY 1152 EKESQPEKLOKLPKFAHSHLMOHELNRSSKSSASEK 1188
Db 1258 PKDKKPPPVSKQKMNAGTLNLTN---PLSNIGSSKQK 1291

```

```

RESULT 11
ZAP3_HUMAN
ID ZAP3_HUMAN STANDARD; PRT; 1822 AA.
AC P49750; P49752; Q9P1V7;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear protein ZAP3 (ZAP13).
GN ZAP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloom S., Dors M., Dichtoff R., Fleetwood P., Harrison G., James R.,
RA Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.;
RT "Sequencing of human chromosome 14q24.3 region."
RL Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95319502; PubMed=7596406;
RA Sherrington R., Rogaev E.T., Liang Y., Rogaeva E.A., Levesque G.,
RA Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,
RA Foncin J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rainero I.,
RA Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A.,
RA Sausseu P., Polinsky R.J., Masco W., da Silva H.A.R., Haines J.L.,
RA Petrick-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
RA Rommens J.M., St George-Hyslop P.H.;
RT "Cloning of a gene bearing missense mutations in early-onset familial
RT Alzheimer's disease."
RL Nature 375:754-760(1995).
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1661.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AC007956; AF61275.1; -
DR EMBL: L40403; AAC42008.1; ALT_FRAME.
DR EMBL: L40400; AAC42006.1; -.
KW Nuclear protein.
FT DOMAIN 15 205 PRO-RICH.
FT DOMAIN 382 430 GLN-RICH.
FT DOMAIN 807 1209 ARG-RICH.
FT DOMAIN 1488 1577 ARG-RICH.
FT CONFLICT 621 621 P->S (IN REF. 2).
FT CONFLICT 1404 1404 T->I (IN REF. 2).
FT CONFLICT 1821 1821 K->E (IN REF. 2).
SQ SEQUENCE 1822 AA; 204947 MW; 8E6CB83FE240C7D2 CRC64;

```

Query Match 4.1%; Score 305; DB 1; Length 1822;
 Best Local Similarity 19.4%; Pred. No. 7.9e-05;
 Matches 333; Conservative 192; Mismatches 554; Indels 634; Gaps 84;

```

QY 139 PPLFLVPPPP-LSQRLQPRSRPILRQD---PVYFRVTVPSTGCDTITAGSC---S 191
Db 102 PPPPPPPPPALSYQKQDQYKQMLHNRDGGPLVPMLESPSPVPVPPSYMPS 161
QY 192 DSGTLRPLASISISTNOSGLAPVGNPSSAPAPAPDITATVYSIVHSEKSDGPPRD 251
Db 162 QSYMPPPPPPSSTYPPSSQPYL-PPAQPSQSPSPSSQSYLAPTPSYSSSSSSQSYLSH 220

```


| | | | | | | | | | | | | | | | |
|----|------|------------------|----------------|----------------|------------------|-------------|----------------|---------------|----------|---------|---------|-------|--------|---------|------|
| OY | 252 | SDNKLRLPAALTSNHS | GASISALIT | ALMEKGVPR | ----- | SPMWSIFV | GOSLHGQI | 305 | | | | | | | |
| Db | 221 | SOSTLPSQASPSRPS | QSHSLSQILAP | PPPSAPRPGKTTVQ | QOAPRLPSGAKN | STBOQDA | 280 | | | | | | | | |
| OY | 306 | PTTGPVRLNARPCG | GGPGW | -EH | ----- | SNKLGIYS | -PPQIRRGESRC | 348 | | | | | | | |
| Db | 281 | PEPDP | ----- | STWTPOEOQYVY | ROHLSLSQRTKVHL | PGHKKG | GVYAKDIP | PEVYKEBYV | 336 | | | | | | |
| OY | 349 | YRSINRGHHSESRQ | RSQRTOSL | PLATPEFYV | ----- | PPLYP | ----- | 386 | | | | | | | |
| Db | 337 | ----- | PATQVSPESPS | SEBPP | -LPPNEEVP | PLPPEPQSED | EDDEARLKQI | 384 | | | | | | | |
| OY | 387 | ----- | ----- | ----- | PPH | ----- | TLPLPGVPP | 401 | | | | | | | |
| Db | 385 | AAAAHMOOHOHRG | FPYOGIMQKHQLO | LOOYOOIIO | PPHHIQA | TTP | -PPGLPPRGV | 443 | | | | | | | |
| OY | 402 | ----- | QSPPOPSOPPTAGS | ----- | VPPPGFPAPANISTAC | FSBPVTAHSNT | PTTQA | 455 | | | | | | | |
| Db | 444 | PGQIPOLTLAPV | PASSOSQVPE | ----- | KPRALLPT | ----- | PVSGSAPPTTHPP | 493 | | | | | | | |
| OY | 456 | LSREEFYREQD | KREKFPYSSGYS | RSSTSSQGLAQH | HALTS | SAHITD | LHD | 515 | | | | | | | |
| Db | 494 | LOSAGPSQVY | ----- | SKAPLSKALPYSES | -SDGLES | ----- | SAAPSQITPAVD | 541 | | | | | | | |
| OY | 516 | HPHPPEAEKASAM | IYMPDILMI | IAHARS | -RSBPY | ----- | RRYR | 564 | | | | | | | |
| Db | 542 | MP | ----- | VRSGL | -LDP | PRSSYLES | RGPRFPGPRFEDLS | RGCRGPRKPREG | 592 | | | | | | |
| OY | 565 | QSP | ----- | TKN | ----- | VPREKEKRYFN | RYEVPPIY | IKAYIGRSVDFDP | 607 | | | | | | |
| Db | 593 | NRPGPRREYEGH | PAEGTCKKMG | WIPKGAPO | YITP | STLS | P | 635 | | | | | | | |
| OY | 608 | PEKRYREMEKRYE | WEYKYGGA | VOAPRPSANRE | ----- | DESPERLY | LNINSPT | R | 663 | | | | | | |
| Db | 636 | ----- | RQSGQWKKPRKPA | FQOQHQOQK | QGAEBL | -SGNKEPLAD | SSNQKFKQMSAF | SLA | 692 | | | | | | |
| OY | 664 | GREDYAAQOSHRR | NRNIGNY | PEKLT | STROS | ----- | HAKNPYSKE | KE | 706 | | | | | | |
| Db | 693 | ADVADYKAAOS | NENL | LSQOQEP | KRESEYSE | GPVPSMWD | QVMSMETQID | KAQAVTQVPL | 752 | | | | | | |
| OY | 707 | ----- | ----- | ----- | SENVG | GDCKG | NKHKH | 722 | | | | | | | |
| Db | 753 | ANKVVPASQ | TFPSKGTGME | GTAVAN | SSLADMD | FKPVG | IGLPHSN | -NDQKGLPRDN | 810 | | | | | | |
| OY | 723 | RKRNEKEGBES | EFLNELLET | SRKCRSSG | IDETK | TDLFV | DSRDA | ----- | 772 | | | | | | |
| Db | 811 | RDNLEGNR | NGSSSYRGP | ----- | GOSRMED | YR | -DKGLVNRG | QALISRG | GLVKQ | 859 | | | | | |
| OY | 773 | TPVDEPDMA | ESTIFKVS | QDKREK | ----- | DKPKYSQ | KTRK | SGSATA | TAKKDVL | 824 | | | | | |
| Db | 860 | EDFDMKMMGR | ----- | EDSNEKNNR | EGS | RDLGLV | PGSSSRKQ | PPGLQSG | SQDNGA | 910 | | | | | |
| OY | 825 | KPS | ----- | KGPOE | VGDGRKES | SRSEPL | PKKAEATK | IDSVK | PSSSSQ | QDEKVTG | PRKAKS | 882 | | | |
| Db | 911 | AGSHERGP | PRRA | -GSGRG | ----- | PLRRAGSR | ----- | ERLPR | PRASRER | -GPR | RGGS | 957 | | | |
| OY | 883 | KSAQDTR | ROSOP | TRRSK | FTVPT | SSQKOPV | -RT | RRPS | ----- | LRK | INY | 927 | | | |
| Db | 958 | RE | -GLGRSP | GRDGRG | FRPE | PDGGEKMT | PYHRD | PPAPANNH | GBENGHEE | FF | PLDGRNA | 1016 | | | |
| OY | 928 | LIAREKN | ----- | EREKRS | SVKDESS | SMKISK | VEGE | EIVKPS | ----- | ----- | ----- | 966 | | | |
| Db | 1017 | PMEERLDD | MDREYV | RECERD | YQD | TLEL | YNED | FXSAP | SRSHD | GRG | RWMWMD | ERD | 1076 | | |
| OY | 967 | ----- | PKRMGE | VLEK | LETP | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 991 | | |
| Db | 1077 | QDMEDY | NREMERD | MDVD | YRIS | PMIDMY | DRSLDN | EMWD | RDY | GAP | LDQESQ | R | EDIPSL | 1136 | |
| OY | 992 | TPAKKI | ----- | ----- | KLNETG | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 1004 | | |
| Db | 1137 | PLPLPLPL | PLD | RYRDRD | NRREK | NRHGH | QDRD | FRD | RGEL | RI | REY | PERG | DT | REKRDYV | 1196 |

| | | | | | | |
|---------|---|--|--|----------------|-------------|-------|
| QY | 1005 | -----KTIQGN----- | -AENASTTREPETKLESTNS | 102 | | |
| Db | 1197 | DRDMERERLSDRMYPESDVIDRHSHPMEHMPSHSHSEMGSDASLSDDQGLGAVYLSQR | 12366 | | | |
| QY | 1029 | -----KTKQKRVGKAKRKVAVGEGSSSTLVYD----- | TSTSS----- | 1061 | | |
| Db | 1257 | QHETILKAQGLTKMLRBOQKPDQLKMKPGFSE-- | PQMAHDLPOESRLONTSSRPGWYP | 13133 | | |
| QY | 1062 | -----TGGSVYKRSSEKTEDTDKRYIK-- | TMEIYNNDNTAPAEVYIMIQ-- | 11033 | | |
| Db | 1314 | PGSYRPPPMGKPPGSIVRPSAPARASVVPVTPRPPPPPPPLPPPPVPIKQTS | 13737 | | | |
| QY | 1104 | VPOSAMKDD | -----DFESEEDBVTKT----- | OP10S----- | 11229 | |
| Db | 1374 | VEQEWEDSDSYGLMDPTINDDQGLNSEKSKSTALIPAPLPPPPVPHSSIPPGPVPMQM | 14333 | | | |
| QY | 1130 | -VGKPSIT----- | KNVTPKPSATAKTEKESQPEKLOKLPEKASHLMQHELNSS | 11800 | | |
| Db | 1434 | PMSPRPPVQQTVDYGHGRDRISTNKEVQIPGRKITLRPD-- | PLPERSTFE----- | 14811 | | |
| QY | 1181 | KGSASSSEKRAKDEHSGSKMDPKRKSQAQDKESTYDR---- | LSBQGHKTKTLSQSK | 12366 | | |
| Db | 1482 | -----TEHNG--ORDRYDRER----- | DREPIYDROSNTYADHROK-- | RDR | 1518 | |
| QY | 1237 | ET--RTSEKHESV---- | GGSSNKKDFTPGRKKYVD-- | DSRDYSSSKRRD-- | ERGLARKKDSP | 1289 |
| Db | 1519 | ETHNRDORDREVIDYDPRDRPREPRDDAQSRYRDKHSSSRGGFDRPESYDKSDRP | 15787 | | | |
| QY | 1230 | P-RGKESLISGQSKSLRER-- | DLPKKAESKSNSSPPRCKTHDKAPETKRPCEETKP | 1347 | | |
| Db | 1579 | VEYEGPMFGCGRRRTYPERMPALPAPSL---- | SHQPP----- | APRVEKKP-- | ESKN | 16233 |
| QY | 1348 | VD-----KNSGKE--REKHAAEARNKESSGANC | 1376 | | | |
| Db | 1624 | VDDILKPGRGRESRPERIVYIMRGLPGSG-- | KTHY | 1655 | | |
| RESULT | 12 | | | | | |
| APC_RAT | | STANDARD; | PRT; | 2842 | AA. | |
| AC | P70478; | | | | | |
| DT | 16-OCT-2001 (Rel. 40, Created) | | | | | |
| DT | 16-OCT-2001 (Rel. 40, last sequence update) | | | | | |
| DT | 16-OCT-2001 (Rel. 40, last annotation update) | | | | | |
| DE | Adenomatous polyposis coli protein (Apc protein). | | | | | |
| GN | Apc. | | | | | |
| OS | Rattus norvegicus (Rat). | | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | |
| OC | Mammalia; Eutheria; Rodentia; Muridae; Murinae; Rattus. | | | | | |
| OX | NCBI_TaxID=10116; | | | | | |
| RN | [1] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | |
| RC | STRAIN=FISCHER 344/N; TISSUE=Brain; | | | | | |
| RX | MEDLINE=96116966; Pubmed=8563176; | | | | | |
| RA | Toyota M., Ushijima T., Kakuchi H., Watanabe M., Imai K., Yachi A., | | | | | |
| RA | Sugimura T., Nagao M.; | | | | | |
| RT | "cDNA cloning of the rat Apc gene and assignment to chromosome 18."; | | | | | |
| RL | Mamm. Genome 6:746-748(1993). | | | | | |
| RN | [2] | | | | | |
| RP | MUTAGENESIS. | | | | | |
| RC | STRAIN=SPRAGUE-DAWLEY, AND FISCHER 344/N; | | | | | |
| RX | MEDLINE=95148647; Pubmed=7846077; | | | | | |
| RA | Kakuchi H., Watanabe M., Ushijima T., Toyota M., Imai K., | | | | | |
| RA | Weisburger J.H., Sugimura T., Nagao M.; | | | | | |
| RT | "Specific 5'-GGCA-3'->5'-GGA-3' mutation of the Apc gene in rat colon | | | | | |
| RL | tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine."; | | | | | |
| CC | Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995). | | | | | |
| CC | -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA- | | | | | |
| CC | CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE | | | | | |
| CC | ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATENIN (BY | | | | | |
| CC | STIMULATORY). | | | | | |
| CC | -1- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATENINS (BY | | | | | |
| CC | SIMILARITY). | | | | | |

Db 2616 KESEIPNTWVSGTSSGA 2634

RESULT 13

MAPB_RAT STANDARD; PRT: 2459 AA.
 AC P15205; Q62958; Q9ER21; Q9QW92;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
 light chain LC1].
 GN MAP1B.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN RN
 RP SEQUENCE OF 1-142 FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Testis;
 RX MEDLINE=96257242; PubMed=8666295;
 RA Liu D., Fischer I.;
 RT "Isolation and sequencing of the 5' end of the rat microtubule-
 associated protein (MAP1B)-encoding cDNA.";
 RL Gene 172:307-308(1996).
 RN RN
 RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Brain, and Gli3 tumor;
 RX MEDLINE=92347374; PubMed=1639092;
 RA Zauner W., Kratz J., Staunton J., Felck P., Wiche G.;
 RT "Identification of two distinct microtubule binding domains on
 recombinant rat MAP 1B.";
 RL Eur. J. Cell Biol. 57:66-74(1992).
 RN RN
 RP [3]
 RC SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
 RX TISSUE=spinal cord; PubMed=2555150;
 MEDLINE=90059871; PubMed=2555150;
 RA Rittenz A., Greeningloh G., Hermanns-Borgmeyer I., Kirsch J.,
 RA Litaier U., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
 RT "Neuraxin, a novel putative structural protein of the rat central
 nervous system that is immunologically related to microtubule-
 associated protein 5.";
 RL EMBO J. 8:2879-2888(1989).
 RN RN
 RP [4]
 RC DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
 RX MEDLINE=97405699; PubMed=9260743;
 RA Ma D., Nothias F., Boyne L.J., Fischer I.;
 RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
 in rat CNS and PNS during development.";
 RL J. Neurosci. Res. 49:319-332(1997).
 CC -1- FUNCTION: The function of brain MAPs is essentially unknown.
 CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
 CC that accompany neurite extension. Possibly MAP1B binds to at least
 CC two tubulin subunits in the polymer, and this bridging of subunits
 CC might be involved in nucleating microtubule polymerization and in
 CC stabilizing microtubules.
 CC -1- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
 CC with MAP1A and MAP1B proteins.
 CC -1- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
 CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
 CC heart or muscle.
 CC -1- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
 CC nerve levels are high early in development but decrease during
 CC postnatal development and are low in adults. In dorsal root
 CC ganglia levels remain high throughout development.
 CC -1- INDUCTION: By nerve growth factor.
 CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
 CC KKEE and KKEIV, repeated but not at fixed intervals, which is
 CC responsible for the binding of MAP1B to microtubules.
 CC -1- PFM: ICI is coexpressed with MAP1B. It is a polypeptide generated
 CC from MAP1B by proteolytic processing. It is free to associate with
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region
 CC of MAP1B (by similarity).

CC -1- PFM: Phosphorylated.
 CC -1- SIMILARITY: TO MAP1A.
 CC CAUTION: A C-terminal fragment of this protein (residues 1597 to
 CC 2459) was originally described as neuraxin in ref.3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 CC EMBL: U52950; AB17068.1; -
 CC EMBL: X60370; CAC16162.1; -
 CC EMBL: X16623; CAA34620.1; ALT_SEQ.
 CC PIR: S06017; S06017.
 CC Interpro: IPR000102; MAP1B-neuraxin.
 CC Pfam: PF00414; MAP1B-neuraxin; 10.
 CC PROSITE: PS00230; MAP1B-NEURAXIN; 8.
 CC Microtubules; Repeat; Phosphorylation.
 CC CHAIN ? 2459
 CC REPEAT 1869 1885 MAP1B 1.
 CC REPEAT 1886 1902 MAP1B 2.
 CC REPEAT 1903 1919 MAP1B 3.
 CC REPEAT 1920 1936 MAP1B 4.
 CC REPEAT 1937 1953 MAP1B 5.
 CC REPEAT 1954 1970 MAP1B 6.
 CC REPEAT 1988 2004 MAP1B 7.
 CC REPEAT 2005 2021 MAP1B 8.
 CC REPEAT 2022 2038 MAP1B 9.
 CC REPEAT 2039 2055 MAP1B 10.
 CC REPEAT 559 1035 GUD-RICH.
 CC DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 CC KKEE AND KKEIV REPEATS).
 CC FT DOMAIN 2224 2312 LYS-RICH.
 CC FT REPEAT 127 127 M -> V (IN REF. 1).
 CC FT REPEAT 140 140 T -> S (IN REF. 1).
 CC FT REPEAT 2112 2112 R -> K (IN REF. 3).
 CC FT REPEAT 2169 2169 L -> I (IN REF. 3).
 CC FT REPEAT 2459 2459 AA: 269497 MW: 283687ZDEDB8BA2 CRC64;
 CC SQ SEQUENCE

Query Match 4.1%; Score 299.5; DB 1; Length 2459;
 Best Local Similarity 23.3%; Pred. No. 0.00017;
 Matches 133; Conservative 119; Mismatches 282; Indels 103; Gaps 32;
 QY 801 PKVSDTKRKSGSATK-----KDNVLRPSK--GPQEKVDGKSPRSEPLK-- 849
 DB 527 PPKVQYKLRQSRSLKPKATPKPLSKSVKSKSEKAPATKASQVEKPKVESKEKVI 586
 QY 850 KAKEATKIDSVKPS--SSQKDEKVGTPPKAKSKSKADTRROSOPRTSKRTVPTSS 908
 DB 587 VKKDKPKGVES--KPSYTEKEVPSKEQSPVKA--EVAEKATSPKPVYKDK--VVKKEIK 642
 QY 909 OKQOPVTRPPRS-----LRKINYLAREKNRERKKSVDKDFSSSMKISKYE 958
 DB 643 TKPEKKKEPKKEVAKKEDKTPPKK-----DEKPKKEAKKEIKKEKKEKELK-- 694
 QY 959 GTEIVRSPKRMKGDEVEKLETPPEKDIASSTPPAKKIKLINBETGKIGNAENASTYKE 1018
 DB 695 -KEYKKEPKLAKKKEVKKDEK---KEYKKEKPEKKEIK---KISKDKSPPLSDTK 747
 QY 1019 P-----SEKLESTSK--IKQEKVKGAKRKVAGSEK-----SSSTLVDTSTSSGSGSP 1066
 DB 748 PALAKPKVAKKEPKETKEPIAAGLKDKGKVYKIKKGGTTEAAATVGTAAVAAAGVA 807
 QY 1067 VKRSEKTEPKRVIYKIMEYVND--NPAADVIYIMQV--PQSMWDKDFSESEEDVKT 1124
 DB 808 ASGPAKLEAERSIMSSPELDITDFELKAEEDIVANDIKPQLELDD-----EKKKET 862
 QY 1125 OPIOS--VGRPSSIKKVVTKPSATAKYTEKESK---QPEKLOKLPKASHELMQHELRS 1179
 DB 1125 OPIOS--VGRPSSIKKVVTKPSATAKYTEKESK---QPEKLOKLPKASHELMQHELRS 1179

```

Db 863 EPGELVYVQKTEVNSKGSAAEPDEGITTTEBEGECOEQPELLEPEVKQGVNDIKFDEEG 922
Oy 1160 SKGSASSKSG---RAKREHSGSEKNDPDKRKSQA---QPKDESTYDRLSQGHFKTLS 1232
Db 923 AGFESSSSGAEVYEEKAETEEAEPEDEDEDVSGASKSHSPTEDEBEIKAADVHIKEKR 982
Oy 1233 QS--SKERTSEKHSV--RGSSNMDFRPG---RDKKYDVOSRDYSSSKRR----- 1276
Db 963 ESVAAGDRAEDMDALEKEADESEEGEEBEEDKADAREEHPDKTEADYVMAV 1042
Oy 1277 DERGELARRKD-----SPPR--GKESLSGOKSKLREERDLPKKGAESKKNSSSP--RDK 1327
Db 1043 DKAENAGTDEQDYDELGRPAKQPGVSPSRPASSIHETLP--GGSSEEAFAASDENED 1101
Oy 1328 KPHDHKA---PYTKRPECETKPYVDKNSGKREKHAAEANRGESSGANCAYLTR 1360
Db 1102 QPEFTTASGYQTSTIETISSEPTPEDEMS--PRDVMDETNNETESPQSQFVNITK 1157

RESULT 14
ID ATRX_MOUSE STANDARD; PRT; 2476 AA.
AC Q61687;
Dt 15-JUL-1999 (Rel. 38, Created)
Dt 15-JUL-1999 (Rel. 38, Last sequence update)
Dt 01-MAR-2002 (Rel. 41, Last annotation update)
De Transcriptional regulator ATRX (X-linked nuclear protein)
De (Heterochromatin protein 2) (HPI alpha-interacting protein) (HPI-BP38 protein).
GN ATRX OR XNP OR HP1BP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A..
RX MEDLINE-98213653; PubMed-9545503;
RA Picketts D.J., Tasthan A.O., Higgs D.R., Gibbons R.J.;
RT "Completion of the human and murine ATRX gene identifies highly
RT conserved, functionally important domains.";
RL Mamm. Genome 9:400-403(1996).
RN [2]
RP SEQUENCE OF 325-1176 FROM N.A.
RX MEDLINE-97133299; PubMed-8978696;
RA le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,
RA Jeanmougin F., Lossen R., Chambon P.;
RT "A possible involvement of Trp1 alpha and Trp1 beta in the epigenetic
RT control of transcription by nuclear receptors.";
RL EMBO J. 15:6701-6715(1996).
RN [3]
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP HETEROCHROMATIN.
RX MEDLINE-20040663; PubMed-10570185;
RA McConwell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT "Localization of a putative transcriptional regulator (ATRX) at
RT pericentromeric heterochromatin and the short arms of acrocentric
RT chromosomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
RN [4]
RP FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
RN GENE EXPRESSION BY AFFECTING CHROMATIN.
CC -1- SUBUNIT: PROBABLY BINDS EHZ2. BINDS ANNEKIN V IN A CALCIUM AND
CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYL SERINE-DEPENDENT MANNER (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS. PROBABLY BY
CC INTERACTING WITH HPI.
CC -1- SIMILARITY: BELONGS TO THE SNE2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

[illegible]

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: Z34289; CAB4063.1; -
DR EMBL: D21262; BAA04803.1; -
DR MIM: 602394; -
KW Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding;
KW Alternative splicing.
FT DOMAIN 84 566 11 X 12 AA APPROXIMATE REPEATS OF AN
FT REPEAT 84 95 ACIDIC SERINE CLUSTER
FT REPEAT 125 95 ACIDIC SERINE CLUSTER 1.
FT REPEAT 167 178 ACIDIC SERINE CLUSTER 2.
FT REPEAT 221 232 ACIDIC SERINE CLUSTER 3.
FT REPEAT 264 275 ACIDIC SERINE CLUSTER 4.
FT REPEAT 325 336 ACIDIC SERINE CLUSTER 5.
FT REPEAT 363 375 ACIDIC SERINE CLUSTER 6.
FT REPEAT 425 436 ACIDIC SERINE CLUSTER 7.
FT REPEAT 470 481 ACIDIC SERINE CLUSTER 8.
FT REPEAT 519 529 ACIDIC SERINE CLUSTER 9.
FT REPEAT 555 566 ACIDIC SERINE CLUSTER 10.
FT DOMAIN 68 82 ACIDIC SERINE CLUSTER 11.
FT DOMAIN 204 382 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 384 587 INTERACTS WITH RPA194.
FT DOMAIN 601 617 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 563 563 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT VARSPPLIC 241 241 PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT CONFLICT 3 3 K -> A (IN REF. 2).
FT CONFLICT 133 133 D -> S (IN REF. 2).
FT CONFLICT 291 292 R -> S (IN REF. 2).
FT CONFLICT 456 456 YA -> SV (IN REF. 2).
FT CONFLICT 456 456 S -> P (IN REF. 2).
SQ SEQUENCE .699 AA; 73720 MW; DFD4AD94EDF659FB CRC64;

Query Match 4.0%; Score 294; DB 1: Length 699;

Best Local Similarity 21.2%; Pred. No. 8.1e-05;
Matches 155; Conservative 121; Mismatches 279; Indels 176; Gaps 30;

QY 771 DATPDEPMDEASTIFKSVSDKREKDPVKSDTKRKSDGATAKKNV---LKPS 827
DB 3 DAGIRRVPSDLYPLVGLRDNQLESEVANKPAKATGATQO--DANASSLDIYFWLXSA 61
QY 828 KGPQEKVDGDRKSPSPPEPLKKAKEATKIDSVKPSSSQKDEKVTGTPRAHKSARD 887
DB 62 KVPERRIA-----NGPVAKKAKKASSDS--EDSSEEEVQGPAPK--KAAPV 108
QY 888 TTRQOPRTRRRSKRTVPKTS-----QKSPVPRTR-RPRSLRKINILIAEKNE 935
DB 109 ARVGLPPGKAATAKAKASESSSEESRDDDEEDOKOPVQGVKPPQ-----AKAARA 159
QY 936 REKRKSVKDFESSSMKISVGEETIVKSPRKMEGDEVKLETPPEKOKIASSTP-- 993
DB 160 PKKAKSSDSDSDSS-----EDPEPPKN-----QPKITPVYVKAQTAKPPKP 202
QY 994 ---AKIKILNETGKKIGNAENASTTKEPSEKLESTSSK-IKQEKYKGA-----KR 1041
DB 203 ARAAPRIANGAKAASSSSSSSSSSSSSDSSEKKAATPKKTPVKQVAKAPVKAATTPTR 262
QY 1042 KYAGSEGSSTLVD-----YTSSTSGSGSPVRKS-----EKTDTKRTVYIKTM 1084
DB 263 KSSSSSDSDSDEEOKKPMKNGPGYIYAPPSAPPPKSLGTOPPKRAVEKQOPVESS 322
QY 1085 EEFYNDNTAPAD-----YIMI-----QVPQSKMDKDFESEEDVKTQPI- 1127
DB 323 EDSDESDSSSEEEKKPTKAVYAKATTKPPAKKAESSSDSDSDSDEDEAPSKPAG 382
QY 1128 ---QSYGKPSIITKNVTPSATKY-----TEKESQPEKLOKL- 1164
DB 383 TTKNSNKPDAVTTKSPAVVAPPAAPKOPVGGOKLITRKADSSSSSEESSSEEEKTKMY 442

QY 1165 -----PKASHHELM-----OHELRSKGSASSEKRAKDRHSGSEKNDPKRRKSGAO 1212
DB 443 ATTTPRAATKAALSLPAKQAPQSGSDSDSDSSSEEEKTSKSAVKKPQVAGGAA 502
QY 1213 PDKESTVDRLSPQGHFKTLTSSQSKETRTSEKHESVRG-----SSNKDFTPGDRKK 1262
DB 503 PSKPAS---AKKGAESSNSSSDSDSSEEEKIKLGKSPRPQAPKANGTALTAQNGX 558
QY 1263 VDIYSRDYSSSKRR-----DERGELARRK-DSPPRKESLSGQSKLREERDLPRKGAES 1316
DB 559 AAKNSEEEEEEEKKAAYVVKSGSLKRRKQNEAAKEATPQAKIKILOTPNTEPPKR-KKG 617
QY 1317 KKSNSPPR-----DKPHDKAPYETRRPC-----EETKPYDK-NSGKEKEKHA 1361
DB 618 EKRASSPFRVRVEETIYDSRVADNS--FDARRGAAGDWGERANDLFTKGSFRHEXT 675
QY 1362 EARNKGESSGA 1372
DB 676 KKKRGSYRGGS 686

Search completed: September 12, 2002, 19:49:53
Job time: 7233 sec

```

166 GGGATAGTGTCCCTCCAGGGTTCTCCAGCTCCTGCCAATTATTC 215
432 rTtHAlaCySePheSerProGluValProThrAlaHisSerAspThrMetP 449
216 AACCACTTGGTATCATCAGAGTGCAGACAGCTCATTCAAATACCAATCC 265
449 rOThrThrGlnAlaProLeuLeuSerArgGluGluPheTyrArgGluGln 465
266 CAACAACAACAAGCACACCTTTGTCCAGGAGAAATCTATAGAGACAG 315
466 AsnAspLysGluArgGluSerLysPheProTyrSerGlySerTyrSe 482
316 CGACGACTAAAGAGAGTCTAAATCTCCATATGATGTTCTTCTGATATTC 365
482 rArgSerSerTyrThrAspSerSerGlnGlyLeuAlaGlnHisIleHis 498
366 AAGAAGTTCAATATCTTATCTTAATCAAGATCTGTTCAACACGTTTCAC 415
499 AlaLeuThrLeuSerProSerAlaAlaHisThrLeuAspLeuHisAs 515
416 GCTCTTATTCGATCATTCACGCGCTCACATTCCTCTTCTATTCACGG 465
515 PHisProHisProGluGluGluAlaGluAlaArgSerAlaMetIleValH 532
466 TCACCTCCATACCCACAGAGCGAGCGCAAGCCGCAATTCGCTTC 515
532 IsMetProAspLeuMetAspIleAlaHisIleAlaArgSerArgSerPro 548
516 ACGGTCTAGATCTCATGTGATATTCAT.CGATCTAGGTCAAGGTCACCCCT 564
549 TyrArgArgTyrArgSerArgSerArgSerProGluPheArgGlyG 565
565 TACAGACGCTATCATCTCAGATCAAGATCTCCCAACGTTTAGGGACA 614
565 nSerProThrLysArgAsnValProArgGluGluLysGluArgGluLyrP 582
615 GTCTCCATAATTAACGTATATCTCAAGGGGAAACAGAACGTAAATATP 664
582 heAsnArgTyrArgGluValProProProTyrAspIleLysAlaTyrTyr 598
665 TTATATATATTCAGAGAAAGTTCACACCATATGATGATGAAGACATATAT 714
599 GlyArgSerValAspPheArgAspProPheGluLysGluArgTyrArgG 615
715 GGGAGAGTGTGACTTAGAGACCCATTGAAAGAAACGCTACCGAGA 764
615 uTPrgLuarGlySerTyrArgGluTyrPyrGluLysTyrTyrLysGlyTyrA 632
765 ATGGGAGAGAAATATAGAGAGTGTATGAAAAATATTATTAAGGTTATG 814
632 laValGlyAlaGlnProArgProSerAlaAsnArgGluAspPheSerPro 648
815 CTGCTGGAGCAGCAGCCTAGACCTCAGCAAAATAGAGAAAGATTTCCTCA 864
649 GluArgLeuLeuProLeuAsnIleArgAsnSerProPheThrArgGlyAr 665
865 GAGAGATTGTTGGCCATTACATCAGCAATCTCTCCCTTCAACAAGAGCGG 914
665 gArgGluAspTyrAlaIaGlyGlnSerHisArgAsnArgAsnLeuGlyG 682
915 CAGGAGAACTATGTGTGGGCAAACTCATAGAAAGTCGAAACATATAGTA 964
682 LysAsnTyrProGluLysLeuSerThrArgAspSerHisAsnAlaLysAsp 698
965 GCAACTATCCAGAAAAGCTTTCAGCAAGAGATGTGTCACAAATCAACAAGAT 1014
699 AsnProLysSerLysGluLysGluSerGluAsnValProGlyAspGlyLys 715
1015 AATCAACAAGTCAAAAGAGAGAGATGTAACCTCCAGAGAGATGTGTA 1064
715 sGlyAsnLysHisLysLysHisArgLysArgArgAsnGluLysGlyG 732

```

```

1065 AGGAATATAGCATTAAGAACACAGAAAAAGAGA.....AAAGGG 1105
732 LuGluSerGluSerPheLeuAsnProGluLeuLeuGluThrSerArgLys 748
1106 AGGAAGTGAAGGTTTCTGACCCAGAGATTATTAGACTTTCAGGAAA 1155
749 CysArgGlySerSerGlyIleAspGluThrLysThrAspThrLeuPheVa 765
1156 TCAGAGAACCTACAGCTGTGAAGAAATTAACACACATCATGTGTGT 1205
765 lLeuProSerArgAspAspAlaThrProValArgAspGluProMetIleAsp 782
1206 TCTCCCAAGTAGAGATGATGCCACCTGTTAGAGATGACCAATGGATG 1255
782 laGluSerIleThrPheLysSerValSerAspLysAspLysArgGluLys 798
1256 CAGATATCATCTTTTAATCATGTGTGAAAAAGAACAGAGAGAAAG 1305
799 AspLysProLysValLysSerAspLysThrLysArgLysSerAspGlySe 815
1306 GATTAACCAAAAGCAAGGCTGATTAACCAAGCAAGCAATGATGATG 1355
815 rAlaThrAlaLysLysAspAsnValLysLysProSerLysGlyProGln 832
1356 TGCTGTCCAAAAAGAAATATTTGTAACCTGTCTAAAGGACCCCAAG 1405
832 LuLysValAspGlyAspArgGluLysSerProArgSerGluProLeu 848
1406 AAAAAGTAGATGAGAAACGTGAGAGATCTCTCGATCTGAACCTCAAT 1455
849 LysLysAlaLysGluGluAlaThrLysIleAspSerValLysProSerSe 865
1456 AAAAAGCAAAAGAGAGACTCCGAGACGTGACAAATCTAAATCATATC 1505
865 rSerSerGluLysAspGluLysValThrGlyThrProArgLysAlaHis 882
1506 TTCTCTCAGAGATGAAAAATCATCTGAGAACCCCGAGAAAACCTCACT 1555
882 eLysSerAlaLysAsp.ThrArgArgGlnSerGlnProArgThrArgAr 898
1556 CTAAATCAGCAAAAGAACCAACCAAGAAACCAAGCAAAAGAGAA 1605
898 gSerLysArgThrValProLysThrSerSerGlnLysSerGlnProValA 915
1606 GTGAAGAGAGCATTTCCAAAGATGCAAAATCAGAAAAGCTAACACTAA 1655
915 rGThrArgArgProArgSerLeu...ArgLysIleAsnTyrLeuIleAla 930
1656 GGAAGAAAGGCCAAGAACCTTAATGAGAAAAACAAACCACTTATATA 1705
931 ArgGluLysAsnGluArgGluLysArgLysLys 941
1706 AGGAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1738

seq_name: gb_est1:A1528817
seq_documentation_block:
LOCUS A1528817 828 bp mRNA linear EST 16-APR-1999
DEFINITION ms13908.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:606878 3', similar to TR:P97868 P97868 RETINOBLASTOMA BINDING
PROTEIN 6 ;, mRNA sequence.
ACCESSION A1528817
VERSION A1528817.1 GI:4442952
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 828)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,

```



```

/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH108"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Salt;
Site:2: NCI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by life technologies. Investigators
providing samples: lothar Hennighausen/Robin Humphreys,
NH"
BASE COUNT      253 a      178 c      213 g      115 t
ORIGIN

```

```

alignment_scores:
  Quality: 1238.00      Length: 249
  Ratio: 5.012          Gaps: 2
  Percent Similarity: 99.197      Percent Identity: 98.394

```

```

alignment_block:
US-09-811-045A-1 x BI453873 ..

```

```

Align seg 1/1 to: BI453873 from: 1 to: 759

```

```

1133 ProSerSerIleIleLysAsnValThrThrLysProSerAlaThrAlaLys 1149
|||||
3 CCATCGAGTATATATAAAATGTCACTACTAAGCCATCGCTACGGCTAA 52

1149 sTyrThrGluLysGluSerGluGluProGluLysLeuGluLysLeuProL 1166
|||||
53 GTACACCCGAGAGAGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 102

1166 ySGLuAlaSerHisGluLeuMetGluHisGluLeuArgSerSerLysGly 1182
|||||
103 AGGAGGCGAGCCAGCGATGATGACAGAGAGAGAGAGAGAGAGAGAGAGAG 152

1183 SerAlaSerSerGluLysGluArgAlaLysAspArgGluHisSerGlySe 1199
|||||
153 AGTGGCTCCAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 202

1199 rGluLysAspAsnProAspLysArgLysSerGluLysGluLysGluLysG 1216
|||||
203 GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 252

1216 lUSeThrValAspArgLeuSerGluGluGlnGlnHisPheLysThrLeuSer 1232
|||||
253 AGAGACACGTGGAGCCGCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 302

1233 GluSerSerLysGluThrArgThrSerGluLysHisGluSerValArgG 1249
|||||
303 CAGTCTTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 352

1249 ySerSerAsnLysAspPheThrProGluArgAspLysLysValAspTyrA 1266
|||||
353 TTCCTCAATATAAGACTCTCCTCTGTTAGAGAGAGAGAGAGAGAGAGAG 402

1266 sPSeArgAspTyrSerSerSerSerLysArgArgAspGluArgGlyLysLeu 1282
|||||
403 ACAGAGAGGATTAATTCAGTTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 452

1283 AlaArgArgLysAspSerProProArgGlyLysGluSerLeuSerGlyG 1299
|||||
453 GCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 502

1299 nLysSerLysLeuArgGluGluArgAspLeuProLysGlyAlaGluS 1316
|||||
503 GAAAAAGCAAGCTGAGAGAGAGAGAGAGATTACCTAAAAAGGGGCGAGT 552

1316 eTylLys..SerAsnSerSerProProArgAspLysLysProHisAspH 1332
|||||
553 CAAACAAACAAGTAATTTCTAGCCCCCAAGAGAGAGAGAGAGAGAGAGAG 602

1332 lSLysAlaProTyrGluThrLysArgProGlyGluGluThrLysProVal 1348
|||||

```

```

603 ATMAAGCCCCCTACGAAACCTAAACGCCCATGTGAAGAGAGAGAGAGAG 652
|||||
1349 AspLysAsnSerGlyLysGluArgGluLysHisAlaAlaGluAlaArgAs 1365
|||||
653 GATTAATAACTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 702

1365 nGlyLysGluSerSerGlyAlaAsnGlyHisValTyrLeuThr 1379
|||||
703 TCGGAAGAGTCCAGTGGTGGCAAAATGCCATGTACTACTAAC 746

seq_name: gb_csl:AU124466

```

```

seq_documentation_block:
LOCUS      AU124466      832 bp      mRNA      linear      EST 23-OCT-2000
DEFINITION AU124466 NT2RM4 Homo sapiens cDNA clone NT2RM400050 5', mRNA
sequence.
ACCESSION  AU124466
VERSION    AU124466.1 GI:10949182
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 832)
            Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
            HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
            ,Y., Sugano,S., Isogai,T.)
            Unpublished (2000)
            Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
            Location/Qualifiers
            1..832
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="NT2RM400050"
            /clone_lib="NT2RM4"
            /cell_type="teratocarcinoma"
            /cell_line="NT2"
            /note="vector: pME18SFL3; mRNA from uninduced NT2 neuronal
            precursor cells"

```

```

BASE COUNT      261 a      199 c      144 g      225 t      3 others
ORIGIN

```

```

alignment_scores:
  Quality: 1152.00      Length: 278
  Ratio: 4.760          Gaps: 3
  Percent Similarity: 87.050      Percent Identity: 83.453

```

```

alignment_block:
US-09-811-045A-1 x AU124466 ..

```

```

Align seg 1/1 to: AU124466 from: 1 to: 832

```

```

1 MetMetGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsnTh 17
|||||
53 ATGATGGAAGGAAAGATCTTAATATGAAGGTGCAATGCTTACCAACAC 102

17 rGlyLysTyrAlaIleProThrIleAspAlaGluAlaTyrAlaIleGlyL 34
|||||
103 TGGAAATATGCAATACCACTATATGATGACAGAGAGAGAGAGAGAGAGAG 152

```


seq. documentation_block: 825 bp mRNA linear EST 23-OCT-2000
LOCUS AU124612 NT2RHMA Homo sapiens cDNA clone NT2RH4/000274 5', mRNA
DEFINITION sequence.

| | |
|--|--|
| ACCESSION | AU124612 |
| VERSION | AU124612.1 |
| KEYWORDS | GI:10949328 |
| SOURCE | EST. |
| ORGANISM | human. |
| REFERENCE | Human sapiens |
| AUTHORS | Eukaryota: Metazoa; Chordata; Craniata; Vertebrata: Euteleostomi; Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo. 1 (bases 1 to 825) |
| TITLE | Ota,T., Makamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and Isogai,T. |
| JOURNAL | HRI human cDNA project (Ota,T., Makamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S., Isogai,T.) |
| COMMENT | Unpublished (2000) Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3951 Fax: 81-438-52-3952 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute. Location/Qualifiers 1..825 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="NT2RM4000274" /clone_id="NT2RM4" /cell_type="teratocarcinoma" /cell_line="NT2" /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells" |
| BASE COUNT | 258 a 198 c 143 g 223 t 3 others |
| ORIGIN | |
| alignment_scores: | |
| Quality: | 1118.00 Length: 276 |
| Ratio: | 4.658 Gaps: 1 |
| Percent Similarity: | 86.957 Percent Identity: 83.333 |
| alignment_block: | |
| US-09-811-045A-1 x AU124612 | .. |
| Align seg 1/1 to: AU124612 from: 1 to: 825 | |
| 1 MetMetGluValIlyAspProAsnMetIySgIyAlaMetLeuThrAsnH 17 | |
| 53 ATGATGGAAGTGAAGATCTTAATATGAAAGTGCAATCTTACCAACAC 102 | |
| 17 rGlyIySTyAlaIleProThrIleAspAlaGuaIyAlaIleGlyL 34 | |
| 103 TCGAAATATGCAATACCACTATAGATCAAGAAGCATATGCAATTGGAA 152 | |
| 34 yLlySgIlyLysProProPheLeuProGluGluProSerSerSerGlu 50 | |
| 153 AGAAAGAGAAACCTCCCTTCTTACAGAGAGACCACTCTTCTTCAGAA 202 | |
| 51 GluAspAspProIleProAlaGluLeuLeuCySLeulIleCySLysAspI 67 | |
| 203 GAAAGATGATCCATCCAGATGAATTTGTTGTCTCATCTGCAAGATAT 252 | |
| 67 eWerThrAspAlaValIleProCyScySgIyAsnSerSerCysAspG 84 | |
| 253 TATGACTGATGCTGTTGATTCCTGCTGCTGGAACAATGACTGATG 302 | |
| 84 luCySIIeArGThrIleuLeuGluSerAspIyHisThrCysProThr 100 | |
| 303 AATGTATAGAACACACTCTCGAATCAGATGAGACACACATGTCGACG 352 | |

101 CysH1sglnAsnAspValSerProkspAlaLeuIleAlaAsnLysPhe 117
117 uArgIlnAlaValaAsnPhelYsAsnGlunHrgIyTrHlySargL 134
403 ACGACAGCGCTGAATTAACCTTCAAAAATGAACCTGGCTATACAAAAAGAC 452
134 euArgLysGlnLeuProPheLeuPheLeuValProProPheArgPro 150
453 TACGAAAACAGTTACCTCTCCACCAACCCCAATACACCTCCGAGACCA 502
151 LeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuArgL 167
503 CTGATTCACAGAGAACCTACACCTCTGATGAAATCTCCGATATCAAGACA 552
167 nGlnAspProValValPheArgTyTrHValSerProHrcysSerAspT 184
553 ACAAGATCCTCTTATGATTCACAGTCAGATCATCTTCATCAACT..... 592
164 hLysTrHAlaLeuLysCysSerAspSerGlyTrHLeuSerArgLeuPro 200
593CACCA 598
201 AlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaProP 217
599 GCTCCGCTCATATCTTCATTAACCTTCAATCAAGTCTTCCTTGCCCCCTCC 648
217 oValSerGlnAsnProSerSerAlaProAlaProValProAspIleThra 234
649 TGTCGTCTGGAATTCGCGCT.TCTGCTCCACGCTCCTG.CCTGATATTAACG 696
234 laThValSerIleSerValHisSerGlnLysSerAspGlyProPheArg 250
697 CAACGATACCATTCATCAGTTCATCAGAAAATACAGATGACCTTTTCNG 746
251 AspSerAspAsnLysLeuLeuProAlaIleAlaLeuThrSerGlnIleS 267
747 GATTCTGATTAATTAATATTGCCAGCTGCAGCTCTTGATCAGAGCAGCTC 796
267 rLysGlnAlaSerSerIleAlaIleThr 276
797 AAAGGNAACCTCTTNAATGCAATTACC 824
seq_name: gb_est1:BB626905
seq_documentation_block:
LOCUS BB626905 644 bp mRNA linear EST 26-OCT-2001
DEFINITION BB626905 RIKEN full-length enriched, 12 days embryo, embryonic body
between diaphragm region and neck Mus musculus cDNA clone
943004O102.5, mRNA sequence.
ACCESSION BB626905
VERSION BB626905.1 GI:16464746
SOURCE EST.
ORGANISM house mouse.
MUS musculus.
Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 644)
Arakawa,T., Carlinici,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
, Himemoto,K., Hori,F., Iehi,Y., Ito,M., Kawai,J., Kondo,H., Koudeh,M.,
, Koye,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakih,
, Shizata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)

```

1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsic.riken.go.jp,
URL:http://genome.gsic.riken.go.jp/
Carninci,P., Shibata,Y., Hayata,N., Sugahara,Y., Shibata,K., Itoh
,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujikawa,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Matahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
,S., Kawaji,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN Integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiysasa,H., Yamane,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsic.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
SOURCE
Location/Organisms
1. .644
/organism="Mus musculus"
/db_xref="taxon:10090"
/cloned="9430040j02"
/cldone.lib="RIKEN full-length enriched, 12 days embryo,
embryonic body between diaphragm region and neck"
/tissue_type="embryonic body between diaphragm region and
neck"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site_1: Salt; Site_2: BamHI; CDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGAAGAAGATCCACAGAGCTTTTGTGGTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 370.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGGAGAAGATTTCGAGTAATAAATTAATATCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI."
BASE COUNT 215 a 150 c 188 g 91 t
ORIGIN
alignment_scores:
quality: 1111.00 length: 214
Ratio: 5.192 gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-811-045A-1 x BB626905 ..
Align seg 1/1 to: BB626905 from: 1 to: 644
1143 lysPROSEAlaThrAlaLysTYRTMGUlySGluseGlnInProcl 1159
11 AAAGCATCGGCTAACGCCTAAGTACACCGGAAGGAAGGAAAGCAGCACGCCCA 51

```

```

1159 uLysLeuGlnLysLeuProLysGlnLysSerHisGlnLeuMetGlnHisG 1176
|||||
52 GAAGACGACAAAGCTTCCCAAGAGGGGAGCCACGAGCTGATGACAGCAG 101
|||||
1176 LuLeuArgSerSerLysGlySerAlaSerSerGlnLysGlyAlaGlyAlaLys 1192
|||||
102 AGCTCAGGAGCTCTCAAGGCGAGTCCGCTCCAGTGAAGAGGCGCAGGCCAAG 151
|||||
1193 AspArgGlnHisSerSerGlySerGlnLysAspAsnProAspLysArgLysSe 1209
|||||
152 GACCGGAGAGACTCAGGGTGGAGAAAGACACCTTACACAGAGAGAGAG 201
|||||
1209 rGlyAlaGlnProAspLysGlnSerThrValAspArgLeuSerGlnG 1226
|||||
202 CGGTGCCACGACAGACAAAGAGAGCACTGTGACCCGCTGAGTGAGCAGC 251
|||||
1226 LysHisPheLysThrLeuSerGlnSerSerLysGlnThrArgThrSerGln 1242
|||||
252 GACATTTTAAAGACTCTCTCAGTCTTCCAAAGAGACAGGAGACTTCAAG 301
|||||
1243 LysHisGlnSerValArgGlySerSerAsnLysAspPheThrProGlyArg 1259
|||||
302 AACGACAGAGCTGTCTGGTGGTCTCTCAATTAAGACTTCACTCTGCTAG 351
|||||
1259 gaSPlysLysValAspTyrAspSerArgAspTyrSerSerSerLysArg 1276
|||||
352 AGACAAAGAAAGTGAGCTACGACAGCAGGATTAATTCAGTTCCAAAGCAA 401
|||||
1276 rgaSPGluArgGlyGlnLeuAlaArgArgLysAspSerProProArgGly 1292
|||||
402 GAGACGAGAGAGTGAAATAGCAAGAGAAAGAAAGCTCTCCCGGGGGC 451
|||||
1293 LysGlnSerLeuSerGlyGlnLysSerLysLeuArgGlnGluArgAspLe 1309
|||||
452 AAAGAGCTCTGTCTGGCGAGAAAGCAAGGTGAGGAGAGAGAGATTT 501
|||||
1309 uProLysLysGlyAlaGlnSerLysLysSerAsnSerSerProArg 1326
|||||
502 ACCTAAAGAGGGGGCGAGTCAGTCAAAAAAGTAATTCAGCCCGCCAGAG 551
|||||
1326 spLysLysProHisAspHisLysAlaProTyrGlnThrLysArgProCys 1342
|||||
552 ACAAAAGCCCATGATCATTAAGCCCCCTTACGAACTTAAAGCCCATGT 601
|||||
1343 GlnGlnThrLysProValAspLysAsnSerGlyLysGlnArg 1356
|||||
602 GAAGAGCAAAAGCCTGATGATTAAGTAACTCTGGAAGAGCGG 643
|||||
seq_name: gb_est2:BG575640
seq_documentation_block:
LOCUS BG575640 821 bp mRNA linear EST 10-APR-2001
DEFINITION 602598704F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4707481 5',
mRNA sequence.
ACCESSION BG575640
VERSION BG575640.1 GI:13583293
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 821)
NIH-MGC http://mhc.nci.nih.gov/.
Mammalian Gene Collection (MGC)
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

```

```

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M10577 row: g column: 02
High quality sequence stop: 802.
FEATURES
    source
        1..821
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4707481"
            /clone_lib="NIH_MGC_87"
            /tissue_type="mammary adenocarcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: breast; Vector: pCMV-Sport6; Site: 1: Not;
            Site: 2: Salt; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.383 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH-MGC Library."
BASE COUNT      206 a      252 c      156 g      205 t      2 others
ORIGIN
alignment_scores:
    Quality: 1100.00      Length: 285
    Ratio: 4.583      Gaps: 3
    Percent Similarity: 84.211      Percent Identity: 79.649
alignment_block:
US-09-811-045A-1 x BG575640 ..
Align seg 1/1 to: BG575640 from: 1 to: 821
134 LeuArgGlyGlnLeuProProPheLeuPheLeuValProProArgPro 150
|||||
24 GTCCGAAACAGTTACTCTCTCCACACACCCCAATTCACCTCGAGAC 73
|||||
150 oLeuSerGlnArgAsnLeuGlnProArgSerArgSerProLysLeuArg 167
|||||
74 ACTGATTCAGAGAGAACCTTACACTCTGATGAGATCTCCGATATCAGAC 123
|||||
167 lngLAspProValAlaPheArgTyrThrValSerProThrCysSerAsp 183
|||||
124 AACAAGATCCTTATGATTCAGTACAGTCAATTCATCACT..... 164
|||||
184 ThrLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPro 200
|||||
165 .....CACCC 169
|||||
200 oAlaProSerLysSerSerLeuThrSerAsnGlnSerSerLeuAlaPro 217
|||||
170 AGTCGCTATATCTTCACTTAATCACTTCAATCACTCTCTGCGCCCTC 219
|||||
217 roValSerGlyAsnProSerSerAlaProAlaProValProAspLysThr 233
|||||
220 CTGTGCTCGAATATCCGCTCTCTCAGCTCCTGATGATTAATCACT 269
|||||
234 AlaThrValSerLysSerValHisSerGlnLysSerSpGlyProPheArg 250
|||||
270 GCAACAGTATCCATATCAGTTCAATTCAGAAATCAAGATGAGACCTTTCG 319
|||||
250 gaSPSerAspAsnLysLeuLeuProAlaAlaLeuThrSerGlnHis 267
|||||
320 GGATTCGATATATAATATGCGCAGCTGCGAGCTTGCATCAGAGCACT 369
|||||
267 erLysGlyAlaSerSerLysAlaLeuThrAlaLeuMetGlnGlnLysGly 283
|||||
370 CAAGGAGACCTCCTCAATTCATACCGCTTATGAGAGAGAGAGAGGT 419
|||||
284 ValProGlyThrSerProThrAsnSerLysPheValGlyLysLeuLeu 300
|||||
420 TA..CCAGTGCCTGCTTGTGGAACCCCATCTTGTGGAAGATCATTAAT 468
|||||
300 euHisGlyGlnLeuLeuProThrGlyProValArgLysAlaAla 316
|||||

```

469 TGCATGACAGTTGATCCCACTGCTCCAGTAAAGATTAATTAATCTGCT 518
 317 ATGProGlyGlyArgProGlyTyrGluHisSerAsnLysLeuGlyTyr 333
 519 CCGTCAGAGGTGGTGTGACACGAGCTGGAGCAATTCACAACTGGCTTA 568
 333 TLeuValSerProGlnGlnLeuArgArgGlyLysArgSerCysTyrA 350
 569 TGTGGTTTCCAGCAACAAATTTAGAAAGAGGAGAGAGGAGCTGCTACA 618
 350 rSerLeuAsnArgGlyArgHisSerGluArgSerGlnArgThrGln 366
 619 GAGGTAAACCGTCGGCCGACACAGCAAGCATCACAGAGACTCA 668
 367 SerProSerLeuProAlaThrProCysPheValProValProProPro 383
 669 GGGCCCGCATACACAGCAACTCCAGTCTTGTACTGTCTCCACACT. CC 717
 383 oLeuTyrProProProHisThrLeuProLeuProProGlyValPro 400
 718 TTTGTATCCGCTCCCTCCATACACTCTCCCTCCGCGGTGCTCTC 767
 400 roProGlnPheSerProGlnPheProSerSerGlnProProThrAlaG 416
 768 CTCACACAGTTTCTCTCTCAGTTTCTCCCTGCGCAGACACACCGCTGG 817
 416 y 416
 818 T 818

seq_name: gb_est2:BI686114

seq_documentation_block:

LOCUS BI686114 888 bp mRNA linear EST 18-SEP-2001
 DEFINITION 60313103F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5352960 5',
 mRNA sequence.

ACCESSION BI686114.1 GI:15648742

VERSION EST.

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 888)

AUTHORS NIH-MGC

TITLE NIH-MGC

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contract: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1A11897 row: f column: 01

High quality sequence stop: 757.

FEATURES

source

1..888

Location/Qualifiers

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5352960"

/clone_1ib="NCI_CGAP_Mam6"

/sex="female, virgin"

/tissue.type="infiltrating ductal carcinoma"

/dev_stage="5 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 211 a 183 c 255 g 239 t
 ORIGIN

alignment_scores:

Quality: 1080.00

Ratio: 4.320

Percent Similarity: 80.386

Percent Identity: 74.277

alignment_block:

US-09-811-045a-1 x BI686114/rev ..

Align seg 1/1 to reverse of: BI686114 from: 1 to: 888

74 TLeProCysCysGlyAsnSerSerCysAspGluCysIleArgThrThrLe 90
 880 ATTCCCTGCTGTGGAGACAGTAGCTGTGATGAAGGTAGAGACAGCAT 831
 90 uLeuGluSerAspLysHisThrCysProThrCysHisGlnAsnAspVal 107
 830 CTGGAGCCAGATGACAAACAGGTCCACACAGTGCACCAATGATGTTCT 781
 107 erProAspAlaLeuIleAlaAsnLysPheLeuArgGlnAlaValAsn 123
 780 CCTCG...ATGCTTTACTGCCACAGTTTTCAGCAGCGCTGGTAATAC 734
 124 PheLysAsnGluThrGly.TyrThrLysArgLeuArgLysGlnLeuPro 140
 733 TTTAAATATGAACCTGCTATAC...AGACGACTACGAACACGTTACCTC 687
 140 roPheLeuPheLeuValProProProArgProLeuSerGlnArgAsnLeu 156
 686 CACCCACCCACCCAGTACACACCAAGACCACTACGACGAGGAGACTG 637
 157 GlnProArgSerArgSerProIleLeuArgGlnGlnAspProValAlp 173
 636 CAGCTCTGATGAGATCTCCCATATCAAGACAGAGGATCTCTGATGAT 587
 173 eArgTyrThrValSerProThrCysSerAspThrLysThrAlaGlySer 190
 586 TCCAGTGACATCGCTCCGCTCAGCTCA..... 559
 190 ySerAspSerGlyThrLeuSerArgLeuProAlaProSerIleSerSer 206
 558GTCCTCTATATCTTCA 541
 207 LeuThrSerAsnGlnSerSerLeuAlaProProValSerGlyAsnProSe 223
 540 GTTAACTTCAATCCATCTGCTGCTCTCTCTCTGGAATTCATC 491
 223 rSerAlaProAlaProValProAspIleThrAlaThrValSerIleSer 240
 490 TTTGCCCCAGCTCCAGTACCTGATATACGCAACCGTCTCTATACAG 441
 240 aHisSerGluLysSerAspGlyProPheArgAspSerAspAsnLeu 256
 440 TCCACTCAGAAAAATCGATGAGACCTTTGCGGATTCGATAAATA 391
 257 LeuProAlaAlaAlaLeuThrSerGluHisSerLysGlyAlaSerSer 273
 390 TTGGCAGCTGCGGCCCTTACATCGAACAATTCAAGAGAGAGCTTTCAT 341
 273 eAlaIleThrAlaLeuMetGluLysGlyValProGlyThrSerProT 290
 340 TGTATTTACTGCTCTTATGGAAGAAAGGGGTA.CCAGGTCCAGTCTT 292
 290 rPAsnSerIlePheVal.GlyGlnSerLeuLeuHisGlyGlnLeuIleP 306
 291 GGAACCTCATCTTTGTTGGACAGTCAATTATTCATGACAGCTGATTC 242
 306 oThrThrGlyProValArgIleAsnAlaAlaArgProGlyGlyLysArg 323
 241 CACAACCTGCGCCACTAAGAAATCATGCTGCTGCTGCGTCCAGGTGGCGGC 192

```
333  roglYrPgluHisSerAnlysLeuGlyTyrLeuValSerProProGln 339
|||||
191  CAGCGTGGAGCATTCACACAGCTGGGTACTGATTCTCCACACAG 142
|||||
340  GlnIleAArgArgGlyGluArgSerCysTyrArgSerIleAsnArgGlyAr 356
|||||
141  CAAATTAGAGAGAGAAAAGAGCTGTACAGAGATATMAACCGCGGCG 92
|||||
356  gHisHisSerGluArgSerGlnArgThrGlnSerProSerLeuProAlaT 373
|||||
91  ACACACACAGCAGCATCATCAGAGACATCAAGCCCATCATTCCAGCAA 42
|||||
373  hrProCysPheValProValProProPro 382
|||||
41  CTCACGCTCTTGTGCCCCGTTCCACCACT 13
|||||

seq_name: gb_est2:BG575299

seq_documentation_block:
LOCUS      BG575299                820 bp    mRNA    linear    EST 10-APR-2001
DEFINITION  602597953F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4706773 5',
mRNA sequence.
ACCESSION  BG575299
VERSION    BG575299.1 GI:13582952
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 820)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgaps-remail.nih.gov
            Tissue Procurement: DCTD/DRP
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: L1AM10575 row: 1 column: 14
            High quality sequence stop: 818.
            Location/Qualifiers
                1..820
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4706773"
                /clone_1lb="NIH_MGC_87"
                /tissue_type="mammary adenocarcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: Nct1;
                Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 1.383 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC library."
BASE COUNT  207 a 252 c 155 g 206 t
ORIGIN

alignment_scores:
    Quality: 1073.00      Length: 282
    Ratio: 4.490          Gaps: 5
    Percent Similarity: 84.752      Percent Identity: 80.142

alignment_block:
US-09-811-045A-1 x BG575299  ..
Align seg 1/1 to: BG575299 from: 1 to: 820

134  LeuArgGlyGlnLeuProPheLeuPheLeuValProProProArgPr 150
```

```

:::|||||
24  GTCCGAAAACAGTTACCTTCACACCACCCCAATACCACTCCGAGACC 73
|||||
150  oLeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuArg 167
|||||
74  ACTGATTACAGAGAACCTACAACTCTGATGAGATCTCGATATATAAGC 123
|||||
167  IngInasProValAlaPheArgTyrThrValSerProThrCysSerasp 183
|||||
124  AACAGATATCTCTTATGATTCAGATGACATCTTCATCAACT..... 164
|||||
184  ThrIstThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPr 200
|||||
165  .....CACCC 169
|||||
200  cAlaProSerIleSerSerLeuThrSerAsnGlnSerLeuAlaProp 217
|||||
170  AGCTCCGCTATATCTTCTTACTTCTTACATCAGTCTTCTTGCCCTC 219
|||||
217  roValSerGlyAsnProSerSerAlaProAlaProValProAspIleThr 233
|||||
220  CTGTGTCGGAATCCGCTTCTGCTCCAGGCTCGTGAATGATATACT 269
|||||
234  AlaThrValSerIleSerValHisSerGluysSerAspGlyProPheAr 250
|||||
270  GCACAGTATCCATATCATTCATTCAAGAAAATCAGATGAGCTTTTCG 319
|||||
250  gAspSerAspAnlysLeuProAlaAlaAlaLeuThrSerGluHis 267
|||||
320  GGATTCTATATATAAATATGCGCAGCTGCACCTTCTCATACAGGCACT 369
|||||
267  erLySGlyAlaSerSerIleAlaIleThrAlaLeuMetGlnLysGly 283
|||||
370  CAAAGGAACCTCCCTCAATTCGATACCGCTTATGGAAGAGAGGCT 419
|||||
284  ValProGlyThrSerProThrAsnSerIlePhe_ValGlyGlnSerLeuL 300
|||||
420  TA.CCAGGTGCTCTTCTTGGAAACCCCAATCTTGCTGACAGCATATAT 468
|||||
300  euHisGlyGlnLeuLeuProThr_ThrGlyProValArgIleAsnAla 316
|||||
469  TGCATGAGACAGTTGATCCCAACAACTGCTCCAGTAAGATAATATATGC 518
|||||
316  aArgProGlyGlyArgProGlyTyrPgluHisSerAnlysLeuGlyT 333
|||||
519  TCGTCACAGTGTGCTGCACACGAGCTGGGAACATTCCAACAACTGGCT 568
|||||
333  yrLeuValSerProProGlnGlnIleArgArgGlyGluArgSerCysTyr 349
|||||
569  ATCTGTTTCTCCACACACAAATTAAGAGGGAGAGAGAGCTGCTAC 618
|||||
350  ArgSerIleAsnArgGlyArgHisHisSerGluArgSerGlnArgThrGl 366
|||||
619  AGAAGTATATAACCGTGGCGACACACAGAGAAATATCAAGAGACTCA 668
|||||
366  nSerProSerLeuProAlaThrProCysPheValProValProProp 383
|||||
669  AGGCCGTCATACACAGCACTCCAGTCTTGTACCTTTCACACACT.C 717
|||||
383  roLeuTyrProProPro_ProHisThrLeuProLeuProProGlyValPr 399
|||||
718  CTTTGTATCCGCTCCCTCCCATACACTTCTCTCCCTCCGGGGTGTCC 767
|||||
399  o_ProProGlnPheSerProGlnPheProSerSerGln 411
|||||
768  TCCTCCACAGTTTCTCTCAAGTTTCTCTGCTGCGCAG 805
|||||

seq_name: gb_est2:BG622049

seq_documentation_block:
LOCUS      BG622049                847 bp    mRNA    linear    EST 18-APR-2001
DEFINITION  602646655F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4768333 5',
mRNA sequence.
```


found through the I.M.A.G.E. Consortium/LINL at:
 http://image.lnl.gov
 plate: L1AM1348 row: a column: 06
 High quality sequence stop: 768.

FEATURES
 source location/Qualifiers

1..771
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5141261"
 /clone_id="NIH_MGC_12"
 /tissue_type="cervical carcinoma cell line"
 /lab_host="DH10B"
 /note="Organ: cervix; Vector: pCMV-SF0R6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.4 kb. Library prepared by Life
 Technologies."

BASE COUNT 249 a 181 c 136 g 205 t
 ORIGIN

alignment_scores:
 Quality: 1044.00 Length: 248
 Ratio: 4.879 Gaps: 2
 Percent Similarity: 86.290 Percent Identity: 83.871

alignment_block:
 US-09-811-045A-1 x B1333743 ..

Align seg 1/1 to: B1333743 from: 1 to: 771

```

1 MetMetGluValLysAspProAsnMetLysGlyAlaMetLeuThrAsnTh 17
83 ATGAGGAGAGTGAAGATCCATATGAAAGCGCAATCCTTACCAACAC 132
17 TGLYSTYTAIAlaIleProThrIleAspAlaGluAlaTyrAlaIleGlyL 34
133 TGGAAATATGCAATATCCACTATAGATGACAGACATATGCAATTTGGGA 182
34 YSLYSGlulSerProPheLeuProGluGluProSerSerSerSerGlu 50
133 AGAAGAGAAACCTCCCTTACCAAGAGACCATCTCTTCCACAGAA 232
51 GUAASPAPProIleProAlaGluLeuLeuCysLeuIleCysLysAspI 67
233 GAAGATGATCCATATCCAGATGAATGTTGTCTCATCTGACAGATAT 282
67 GetherAspAlaValAlaIleProCysCysGlyAsnSerSerCysAspG 84
283 TATGACTGATGCTGTGTGATTCCTGCTGCTGGAACATGTTACTGTGATG 332
84 LuCysIleAArgThrThrLeuLeuGluSerAspLysHisThrCysProThr 100
333 AATGTATATAGAACACACTCCTCGAATCAGATGACACACATGTCGACG 382
101 CysHisGlnAsnAspValSerProAspAlaLeuIleAlaAsnLysPheLe 117
383 TGTCTCAAAAAGATGTTCTCCATGCTTAAATGGCCAAATTAATTTT 432
117 uArgGlnAlaValAsnAspPheLysAsnGluThrGlyTyrThrLysArgL 134
433 ACGACAGGCTGTAATACTTCAAAAATGAACTGGCTATACAAAAGAC 482
134 euArgLysGlnLeuProPheLeuPheLeuValProProPheArgPro 150
483 TACGAAAACAGTTACCTCCTCCACACACCCCAATACACACCTCGAGACA 532
151 LeuSerGlnArgAsnLeuGlnProArgSerArgSerProIleLeuArgL 167
533 CTGATTCAGAGAACCTACAACTCTGATGAGATCTCGATATCAAGACA 582
167 nGlnArgProValPheArgTyrThrValSerProThrCysSerAspT 184
583 ACAAGATCCTTATATGATTCACGATGACATCTTCACTCAACT..... 622

```

184 hrLysThrAlaGlySerCysSerAspSerGlyThrLeuSerArgLeuPro 200
 623CACCA 628
 201 AlaProSerIleSerSerLeuThrSerAsnGlnSer.SerLeuAlaProP 217
 629 GCTCGCTATATATCTTCAATTAATCAATGATCTTTCCTGGCCCTGC 678
 217 rovalSerGlyAsnProSerSerAlaProAlaProValProAspIleThr 233
 679 CTGTCTTGGAAATCCGTTCTTCTGCTCCAGCTCCTGTTACTGATTA 728
 234 AlaThrValSerIleSerValHisSerGluLysSerAspGly 247
 729 GCAACAGTATCCATATCAATTCATTCAGAAAATATCAGATGCA 770

seq_name: gb_est2:BF720766

seq_documentation_block:

LOCUS BF720766 744 bp mRNA linear EST 03-JAN-2001
 DEFINITION mab58d02.y1 Soares.thymus.2NDMT Mus musculus CDNA clone
 IMAGE:3974450 5' similar to TR:P70287 P70287 RETINOBLASTOMA BINDING
 PROTEIN 6 ; mRNA sequence.

ACCESSION BF720766 GI:12021768
 VERSION BF720766.1
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 744)
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsb-r@mail.nih.gov
 This clone is available royalty-free through LINL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:1474482

Seq primer: -40RP from Gibco
 High quality sequence stop: 478.

FEATURES
 source location/Qualifiers

1..744
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="3974450"
 /clone_id="Soares_thymus_2NDMT"
 /sex="male"
 /tissue_type="Thymus"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand CDNA
 was primed with a Not I - Oligo(dT) primer [5',
 TGTATCAATCTGAACTGAGAGCGCGCGTGTGTGTGTGTGTGTGTGTGT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through two
 rounds of normalization, and was constructed by Bento
 Soares and M.Fatima Bonaldo."

BASE COUNT 187 a 233 c 138 g 180 t 6 others
 ORIGIN

alignment_scores:
 Quality: 1042.00 Length: 267
 Ratio: 4.673 Gaps: 2
 Percent Similarity: 83.521 Percent Identity: 80.524


```

861 lypsProSerSerSerGlnlysaspGluysValThrGlyThrPro 878
|||||
201 AAAACCTCTCTCTCTTCAGAGAGATGAGAAAGTCACAGAACCCCTA 250
|||||
878 rglYsAlaHISerLySerAlaLysAsp_ThrArgArgGlnSerGlnPr 894
|||||
251 GAAAGCCCACTTCTAAATCTGCAAAAGAACACACGAGGAGCAAGCCAGCC 300
|||||
894 oArgThrArgSerSerLysArgThrValProLysThrSerSerGlnLys 911
|||||
301 AAGGAGAGAGAGGTCAAAAAGAGCTGTTCCAAAGACATCAAGTCAGAAA 350
|||||
911 exGlnProValArgThrArgArgProArgSerLeuArgLysIleasnTy 927
|||||
351 GCCAGCAGTAAGAGAGAGAGAGCCCAAGAGCCTGAGAAAAATAAATACTAC 400
|||||
928 LeuIleAlaArgGluLysasnGluArgGluLysArgLys..LysSerVal 943
|||||
401 TTGATAGCAAGGAGGAGAAAACGAAAGAGAAAACGGAAGAAAGAGTGT 450
|||||
944 AspLysAspPheGluSerSerSerMetLysIleSerLysValGluGlyTh 960
|||||
451 GATRAAGATTGTAGTCGTCTTCAATGAAAATCTCTRAAGTAGAAGAGAAC 500
|||||
960 rgluIleValLysProSerProLysArgLysMetGluGlyAspValGluL 977
|||||
501 AGAAATAGTGAACCATCACTCAATGAAATGGAAGGATGATGTTGAAA 550
|||||
977 ystLeuLysArgThrProGluLysAspLysIleAlaSerSerThrThrPro 993
|||||
551 AGCTGGAAGAGACCCAGTAAGAGACAAAGTTGCATCATCACTACTCTCA 600
|||||
994 AlaLysLysIleLysLeuAsnArgGluThrGlyLysLysIleGlyAsnAl 1010
|||||
601 GCCAAAAAATCAAACTTCACAGAGAAACGCAAAAAAATTTGGAATATGC 650
|||||
1010 acGluAsnAlaSerThrThrLysGluPro 1019
|||||
651 AGAAATGCATCTGTAGTCAAGAACCCCT 678
|||||
seq_name: gb_gss:A2696947
seq_documentation_block:
LOCUS A2696947 782 bp DNA linear GSS 24-JAN-2001
DEFINITION RPCI-23-238D2.TJ RPCI-23 Mus musculus genomic clone RPCI-23-238D2,
DNA sequence.
ACCESSION A2696947
VERSION A2696947.1 GI:12412563
KEYWORDS GSS.
SOURCE
ORGANISM Mus musculus.
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 782)
Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akintret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSS: RPCI-23-238D2.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaod@igf.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pdejong@mai.cho.org). Clones may be purchased from BACPAC
Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
page: http://www.igf.org/ldb/bac_ends/mouse/bac_end_intro.html
Title: JOURNAL
COMMENT
Page: 238 row: D column: 2

```

```

Seq primer: SP6
Class: BAC ends.
FEATURES
    source          location/Qualifiers
                        1..782
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="RPCI-23-238D2"
                        /clone_lib="RPCI-23"
                        /sex="Female"
                        /lab_host="DH10B"
                        /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site-1:
                        EcoRI; Site-2: EcoRI; Female C57BL/6J mouse kidney and/or
                        brain genomic DNA was isolated and partially digested
                        with a combination of EcoRI and EcoRI Methylase. Size
                        selected DNA was cloned into the pBAC3.6 vector at the
                        EcoRI sites. The ligation products were transformed into
                        DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      186 a      145 c      249 g      202 t
ORIGIN
alignment_scores:
    quality: 995.00      Length: 255
    Ratio: 4.502      Gaps: 1
    Percent Similarity: 86.667      Percent Identity: 76.863
alignment_block:
US-09-811-045a-1 x A2696947/rev ..
Align seg 1/1 to reverse of: A2696947 from: 1 to: 782
200 ProAlaProSerIleSerSerLeuThrSerAsnGlnSerSerLeuAlaPr 216
|||||
762 CCACACCCCTCTTAACCTCATTAAGCTTCAAAATTCATCTTCCGCGTCC 713
|||||
216 opProValSerGlyAsnProSerSerAlaProAlaProAlaProAlaPr 233
|||||
712 TTCTGTGCGCTGGAACCCATCTCTCCAGCTCCCGATGCGGAGAACCA 663
|||||
233 hvrAlaThrValSerIleSerValHisSerGluysSeraspGlyProPhe 249
|||||
662 CTCGAAGAGTATCTGTATCAACCTTTCAGAAAACAGATGGAATCCTTT 613
|||||
250 ArgAspSerAspAsnLysLeuProAlaAlaAlaLeuThrSerGlnuH 266
|||||
612 CGGGATTCAGAAAACCAATATTATACAGAGTCCGCCCTTACATCAGAAC 563
|||||
266 sSerLysGlyAlaSerSerIleAlaIleThrAlaLeuMetGluLysG 283
|||||
562 TTCAAAGGAGCGCTTCAATTCCTGTACTCTCTCTATGGAAGAAAGC 513
|||||
283 lylAlaProGlyThrSerProThrPasnSerIlePheAlaGlyInserie 299
|||||
512 GTGG.CCAGGTGGCGGTCCTGAACCTCCACCTTCGTGGACAGTCAC 464
|||||
299 uLeuHisGlyLysLeuIleleProThrThrGlyProValArgLysAlaA 316
|||||
463 ACTGCATGAACAGTATATCCCACTGCTCCAGTAAGATAAAGTCG 414
|||||
316 lAaArgProGlyGlyArgProGlyTyrGlnHisSerAsnLysLeuGly 332
|||||
413 CTCATCCAGTGGTGTCAACAGGTTGGAAACATTCACAAAGAAATGGG 364
|||||
333 TyrLeuValSerProProGlnGlnIleArgArgGlyGluArgSerTy 349
|||||
363 AAGCTGATTCTTCCAAACAGCAAAATTTAGAGTGGAAGAAAGAGCTGTA 314
|||||
349 rArgSerIleAsnArgGlyArgHisSerSerGluArgSerGlnArgTh 366
|||||
313 CCAGAGATTAACAGTGGGAGACATACAGTGAACATCATCAGAGAGAGTC 264
|||||
366 lserProSerLeuProAlaThrProCysPheValProValProProPro 382

```

```
||:::||||| ||||| |||||:::|||||
263 AAGGCCATCAGACGACGAACTCCAGCTTGTGCTATTCACCACT 214
383 ProLeuTyrProProProHisThrLeuProLeuProProGlyValPr 399
||| ||| ||||| ||||| ||||| ||||| |||||
213 CTTTGGCACCCTCTCCACATATGCTTCTCTCTCCAGGTACC 164
399 oProGlnPheSerProGlnPheProSerSerGlnProProThrAlaG 416
| ||:::| ||::: ||::: ||::: ||::: ||::: ||:::
163 TGGCCACAGGTTCTCTCTCGGTTCTCTCGGCGCAGCTCCAAAGCAG 114
416 lYTySerValProProGlyPheProProAlaProAlaAsnIleSer 432
||||| ||||| ||||| ||||| ||||| ||||| |||||
113 GATATAGTGTCCTCTCCAGATTTCACACAGCTCTGCCAATATATCA 64
433 ThrAlaCysPheSerProGlyValProThrAlaHisSerAsnThrMetPr 449
||||| ||||| ||||| ||||| ||||| ||||| |||||
63 ACAGCTTGCTTTCCACCAGGTTCACACTGCCCATTCAAATACCATGCC 14
449 oThrThrGlnAla 453
||||| ||||| |||||
13 CACACACACAGCA 1
```

PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 16:19:12 ; Search time 69.71 Seconds

(without alignments)
3484.219 Million cell updates/sec

Title: US-09-811-045A-1

7374

Sequence: 1 MMEVKDPMKMGALMTWTGK.....WRSSWLIGRWRRAPSSRNPS 1404

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_virus:*
16: SP_bacteriophage:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1 | 6054 | 82.1 | 1591 | 11 | P97868 |
| 2 | 5451.5 | 73.9 | 1616 | 4 | Q96PH3 |
| 3 | 3500.5 | 47.5 | 948 | 4 | Q15290 |
| 4 | 1702 | 23.1 | 628 | 4 | Q9HSM5 |
| 5 | 1305.5 | 17.7 | 529 | 4 | Q9NFX4 |
| 6 | 542 | 7.4 | 1231 | 5 | Q9XZ21 |
| 7 | 480.5 | 6.5 | 1190 | 5 | Q62235 |
| 8 | 405 | 5.5 | 2081 | 10 | Q9LH98 |
| 9 | 396.5 | 5.4 | 2752 | 4 | Q9UQ35 |
| 10 | 396 | 5.4 | 4880 | 11 | Q9ULT1 |
| 11 | 396 | 5.4 | 5085 | 11 | Q9UKS6 |
| 12 | 388.5 | 5.3 | 5120 | 13 | Q9PU36 |
| 13 | 384 | 5.2 | 3484 | 5 | P91257 |
| 14 | 377.5 | 5.1 | 2296 | 4 | Q9UHA8 |
| 15 | 375 | 5.1 | 17352 | 5 | Q95YM2 |
| 16 | 374.5 | 5.1 | 5327 | 5 | Q76891 |

| | | | | | | |
|----|-------|-----|------|----|--------|--------------------|
| 17 | 367.5 | 5.0 | 4833 | 11 | Q9QYX6 | Q9QYX6 mus musculu |
| 18 | 367.5 | 5.0 | 5038 | 11 | Q9QYX7 | Q9QYX7 mus musculu |
| 19 | 363.5 | 4.9 | 6632 | 5 | Q17362 | Q17362 caenorhabdi |
| 20 | 363.5 | 4.9 | 6642 | 5 | Q01761 | Q01761 caenorhabdi |
| 21 | 360.5 | 4.8 | 2083 | 5 | Q9H435 | Q9H435 caenorhabdi |
| 22 | 351 | 4.8 | 2768 | 5 | Q9VC00 | Q9VC00 drosophila |
| 23 | 349 | 4.7 | 990 | 13 | Q91803 | Q91803 xenopus lae |
| 24 | 344.5 | 4.7 | 3111 | 5 | Q9VH10 | Q9VH10 drosophila |
| 25 | 341.5 | 4.6 | 1422 | 6 | Q9SKU4 | Q9SKU4 canis famli |
| 26 | 340 | 4.6 | 5322 | 5 | Q9VPL9 | Q9VPL9 drosophila |
| 27 | 340 | 4.6 | 2703 | 5 | Q9W164 | Q9W164 drosophila |
| 28 | 339 | 4.6 | 2722 | 5 | Q19135 | Q19135 caenorhabdi |
| 29 | 338.5 | 4.6 | 2157 | 11 | Q9Z1R1 | Q9Z1R1 mus musculu |
| 30 | 336.5 | 4.6 | 4498 | 13 | Q93291 | Q93291 fuqu rubrip |
| 31 | 336 | 4.6 | 1062 | 5 | Q960C4 | Q960C4 drosophila |
| 32 | 335.5 | 4.5 | 933 | 4 | Q9UQ14 | Q9UQ14 homo sapien |
| 33 | 335.5 | 4.5 | 1026 | 4 | Q9UJ57 | Q9UJ57 homo sapien |
| 34 | 333 | 4.5 | 1108 | 5 | Q9ND10 | Q9ND10 babesia big |
| 35 | 333 | 4.5 | 1262 | 4 | Q9UQ40 | Q9UQ40 homo sapien |
| 36 | 333 | 4.5 | 1343 | 4 | Q9H7N4 | Q9H7N4 homo sapien |
| 37 | 332.5 | 4.5 | 1151 | 5 | Q9VAY4 | Q9VAY4 drosophila |
| 38 | 331 | 4.5 | 2897 | 5 | Q9VID9 | Q9VID9 drosophila |
| 39 | 331 | 4.5 | 2951 | 5 | Q9W320 | Q9W320 drosophila |
| 40 | 330 | 4.5 | 1430 | 5 | Q9W0B0 | Q9W0B0 drosophila |
| 41 | 329 | 4.5 | 1160 | 11 | Q9ESC8 | Q9ESC8 mus musculu |
| 42 | 328 | 4.4 | 2701 | 4 | Q9Y520 | Q9Y520 homo sapien |
| 43 | 327 | 4.4 | 2187 | 11 | P70670 | P70670 mus musculu |
| 44 | 326.5 | 4.4 | 1341 | 4 | Q9UKV3 | Q9UKV3 homo sapien |
| 45 | 325.5 | 4.4 | 1178 | 5 | Q9NHX5 | Q9NHX5 drosophila |

ALIGNMENTS

RESULT 1
P97868 PRELIMINARY; PRT; 1591 AA.
ID P97868; P70287;
AC P97868; P70287;
DT 01-MAY-1997 (Tremblrel. 03. Created)
DT 01-OCT-2000 (Tremblrel. 15. Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17. Last annotation update)
DE RETINOBLASTOMA BINDING PROTEIN 6 (PROLIFERATION POTENTIAL-RELATED
DE PROTEIN) (P53-ASSOCIATED CELLULAR PROTEIN) (PACT).
GN RBBP6 OR P2P-R.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY (ISOFORM 2).
RC STRAIN-BALB/C;
RX MEDLINE=97188447; PubMed=9037032;
RA Witte M.M., Scott R.E.;
RT "The proliferation potential protein-related (P2P-R) gene with domains
RT encoding heterogeneous nuclear ribonucleoprotein association and Rb1
RT binding shows repressed expression during terminal differentiation.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:1212-1217(1997).
RN [2]
RP REVISIONS.
RC STRAIN-BALB/C;
RX Witte M.M., Scott R.E.;
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 4-1587 FROM N.A., AND CHARACTERIZATION (ISOFORMS 1 AND 2).
RC STRAIN-BALB/C; TISSUE=TESTIS;
RX MEDLINE=97163396; PubMed=9010216;
RA Simons A., Melamed-Bessudo C., Wolkowicz R., Sperling J., Sperling R.,
RA Eisenbach L., Rotter V.;
RT "PACT: cloning and characterization of a cellular p53 binding protein
RT that interacts with Rb1";
RL Oncogene 14:145-155(1997).
CC -!- FUNCTION: BINDS TO Rb1 THROUGH THE Rb1 POCKET DOMAIN. ASSOCIATES
CC WITH HNRNP. ALSO BINDS TO P53. MAY BE INVOLVED IN GROWTH CONTROL

CC -1- AND DIFFERENTIATION. NUCLEAR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS: ISOFORM 1 (SHOWN HERE) AND
CC ISOFORM 2: MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN TESTIS.
CC MODERATE LEVELS FOUND IN HEART, LUNG, LIVER AND SKELETAL MUSCLE.
CC VERY LOW LEVELS DETECTED IN KIDNEY, BRAIN AND SPLEEN. ALSO
CC EXPRESSED IN GROWING UNDIFFERENTIATED MESECHYMAL STEM CELLS BUT
CC NOT DURING THE TERMINAL DIFFERENTIATION PHASE.
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR EMBL: U83913: AAC72432.1; -
DR EMBL: U28789: AAB49620.1; -
DR MGI: MGI:894835: Rdbp6.
DR InterPro: IPR001841: Znf_fing.
DR Pfam: PF00097: zf-C3HC4.1.
DR SMART: SM00184: RING.1.
DR PROSITE: PS00518: ZINC_FINGER_C3HC4_FALSE_NEG.
KW Zinc-finger: Metal-binding; DNA-binding; Nuclear protein;
KW Alternative splicing.
FT ZN_FING 61 101
FT DOMAIN 481 540
FT VARSLIC 1528 1585
FT CONFLICT 454 487
FT CONFLICT 55 55
FT CONFLICT 118 119
FT CONFLICT 142 142
FT CONFLICT 219 219
FT CONFLICT 222 222
FT CONFLICT 381 381
FT CONFLICT 396 396
FT CONFLICT 416 423
FT CONFLICT 430 430
FT CONFLICT 437 437
FT CONFLICT 448 448
FT CONFLICT 490 490
FT CONFLICT 504 504
FT CONFLICT 590 590
FT CONFLICT 742 742
FT CONFLICT 757 759
FT CONFLICT 764 764
FT CONFLICT 779 779
FT CONFLICT 783 783
FT CONFLICT 813 813
FT CONFLICT 1091 1091
FT CONFLICT 1117 1117
FT CONFLICT 1365 1365
FT CONFLICT 1382 1382
FT CONFLICT 1392 1392
FT CONFLICT 1397 1397
FT CONFLICT 1405 1405
SQ SEQUENCE 1591 AA: 177763 MW: 7D0C6798F49D3748 CRC64:

Query Match 82.1% Score 6054; DB 11: Length 1591;
Best Local Similarity 84.7% Pred. No. 1.3e-309;
Matches 1199; Conservative 31; Mismatches 111; Indels 74; Gaps 10;

OY 1 MAEVDPPNKGMLNTGKYAIPITDAEYVAGKKEKPELPEEPESSSEEDDPIPAELL 60
DB 1 MAEVDPPNKGMLNTGKYAIPITDAEYVAGKKEKPELPEEPESSSEEDDPIPAELL 60
OY 61 CLICDINTDAVAVIPCCGSSSCDECIRFTLLSDKHCPTCHQNDVSPALANKFLNQA 120
DB 61 CLICDINTDAVAVIPCCGSSSCDECIRFTLLSDKHCPTCHQNDVSPALANKFLNQA 120
OY 121 VNNFKNETGYTKRLKQKOPPLFLVPPRPRLSQRNLOPRSSPILRODDPVVFRVTSPT 180
DB 121 VNNFKNETGYTKRLKQKOPPLFLVPPRPRLSQRNLOPRSSPILRODDPVVFRVTSPT 180
OY 181 CSDFKTAGSCSDSGLSLRLPAPSSISLTSNOSLAPVSGSPSAPAPVPDITATVSISS 240
DB 181 HS-----APSSISLTSNPSALAPVSGSPSAPAPVPDITATVSISS 222

OY 241 HSEKSDGPPRDSNKLPLAALTSSEKSGASSIATIALMEKNG-----VQSPWNSIFVG 296
DB 223 HSEKSDGPPRDSNKLPLAALTSSEKSGASSIATIALMEKNGVQVPLGTSS-----LLG 278
OY 297 QSLHQQLPTTGQPVYRINARPGGRPGWENHNGKGLYVSPPOQIRGRSCYRSTINCR 356
DB 279 QSLHQQLPTTGQPVYRINARPGGRPGWENHNGKGLYVSPPOQIRGRSCYRSTINCR 338
OY 357 HSEKSORSPSLPATCFVYVPPPLYPPPHNTLPPLPGVPPQSFPPFSSQPPTAG 416
DB 339 HSEKSORSPSLPATCFVYVPPPLYPPPHNTLPPLPGVPPQSFPPFSSQPPTAG 398
OY 417 YSVPPGPPAPANISTACFSQVPTAHNSMTPTQAQLSREBFYRQ----- 465
DB 399 YSVPPGPPAPANISTACFSQVPTAHNSMTPTQAQLSREBFYRQ----- 458
OY 466 -----NDKGN-----SKFYSGSSSYSSSYTDS-SQGLAHIALT 501
DB 459 KLDEFTDFAKELMEYKKIKERRRSRSPKSPSSSYSSSYTDS-SQGLAHIALT 518
OY 502 LSPSAHTLTLHDHPH-----PEAFARSAMVHMDLMDIAHARSPPYRRRSR 557
DB 519 RSPFSRSHSRYSRSPYPRGRGKSRNYSRSH-----GYHRSRSPYRRRSR 573
OY 558 SPPEFRGQSPTRKNVREKEREYFNRYRPPYDIAKAYGSVPDOPFEKERYRME 617
DB 574 SPQAFRGQSPTRKNVREKEREYFNRYRPPYDIAKAYGSVPDOPFEKERYRME 633
OY 618 RYREYKERYGYAVGAOPRPSANREDFPERLLPLNRSNPTGRREDYAAQOSHNN 677
DB 634 RYREYKERYGYAVGAOPRPSANREDFPERLLPLNRSNPTGRREDYAAQOSHNN 693
OY 678 RNLGNYPEKLTSTRSHNANKDPKSKKESENVPCDGGKGNKHKHRRKNEKGESESF 737
DB 694 RNLGNYPEKLTSTRSHNANKDPKSKKESENVPCDGGKGNKHKHRRKNEKGESESF 750
OY 738 LNPelletSRKCRGSGGIDETKDTLFLVPSRDATPVPRDPMNAESTTFYSVSDKRE 797
DB 751 LNPelletSRKCRGSGGIDETKDTLFLVPSRDATPVPRDPMNAESTTFYSVSDKRE 810
OY 798 KDKPRKSDTKRKSDDGATAKKDNVLPKSGPOEKYDGDREKSPRSEPLKAKKEATK 857
DB 811 KDNKVASDKTKRKSDDGATAKKDNVLPKSGPOEKYDGDREKSPRSEPLKAKKEATK 870
OY 858 IDSVKPSSSSQKDEKVTGTPRKAHSAKSAKDRROSQRTKRSKTVKTSOSQOPYRTR 917
DB 871 IDSVKPSSSSQKDEKVTGTPRKAHSAKSAKDRROSQRTKRSKTVKTSOSQOPYRTR 930
OY 918 RPRSLKINYLILAR-EKNEREKRRKSYDXKDFESSSMKISVETGETIYKSPKRMGDVE 976
DB 931 KAKKPEKNNKLDKSGEKRRKRTKESYDQFESSSMKISVETGETIYKSPKRMGDVE 990
OY 977 KLERTEKDIASSTTPAKKIKINRETGKKGNAENASTTKPESEKLESTSSKIKOEKVK 1036
DB 991 KLERTEKDIASSTTPAKKIKINRETGKKGNAENASTTKPESEKLESTSSKIKOEKVK 1050
OY 1037 GAKKRVVAGSESSSTLVDTYSTSGSPVRSKSEKTKTKTVIKNMEYNNDTAPAE 1096
DB 1051 GAKKRVVAGSESSSTLVDTYSTSGSPVRSKSEKTKTKTVIKNMEYNNDTAPAE 1110
OY 1097 DVITIMIOVQSKWMDDEDESEEDVKTQPIOSVKGPSIIKNVTKTSATAKYTEKSE 1156
DB 1111 DVITIMIOVQSKWMDDEDESEEDVKTQPIOSVKGPSIIKNVTKTSATAKYTEKSE 1170
OY 1157 QPEKLOKLPKASHLMOHELRSKSGASSSEKGRAKDRHSGSEKDNDRKSGAOPDKE 1230
DB 1171 QPEKLOKLPKASHLMOHELRSKSGASSSEKGRAKDRHSGSEKDNDRKSGAOPDKE 1246
OY 1217 STVDRLSEOGHFKTLSSSKETRTSEKHEVSGSSNDDFTPGRKVKYDYSDYSSSKRR 1276
DB 1231 STVDRLSEOGHFKTLSSSKETRTSEKHEVSGSSNDDFTPGRKVKYDYSDYSSSKRR 1290
OY 1277 DERGELARRKDSPPRGKESLSGOKSKLREERDLPKKGAESKKNSSPPRDKKHHDKAP 1336

DB 1291 DEGGELARRKDSPPKRESLSGOKSLREERDLPKKAESKSNSSPPDKKHHDHAKAY 1350
 OY 1337 ETKRCEETFPVDKNSGKEREKHAAREARKESG 1371
 DB 1351 ETKRCEETFPVDKNSGKEREKHAAREARKESG 1385

RESULT 2
 OY 096PH3 PRELIMINARY: PRT: 1616 AA.

AC 096PH3: 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE PROLIFERATION POTENTIAL-RELATED PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Camargo A.A., Moreira E.S., Simpson A.J.G.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF352051; AAL05625.1; -;
 SQ SEQUENCE 1616 AA: 182248 MW: DB9F5B4BED95D72B CRC64:

Query Match 73.9% Score 5451.5; DB 4; Length 1616;
 Best Local Similarity 75.3%; Pred. No. 5,6e-278;
 Matches 107; Conservative 92; Mismatches 167; Indels 93; Gaps 16;

OY 1 MMEYDPMKMGAMLTNGKVAIPTIDAEVAIGKKEKPPLEPSSSEEDDPIPAELL 60
 DB 23 MMEYDPMKMGAMLTNGKVAIPTIDAEVAIGKKEKPPLEPSSSEEDDPIPAELL 82
 OY 61 CLICKDITDAVIPCNGSSCDECIRTTTLESKHTPTCHONDVSPDALIANKFLROA 120
 DB 83 CLICKDITDAVIPCNGSSCDECIRTTTLESKHTPTCHONDVSPDALIANKFLROA 142
 OY 121 VNNFNENGYKRLKQKLPFLVPPRPPLSORNLQPSRSPILRQDDPVFRTVSPT 180
 DB 143 VNNFNENGYKRLKQKLPFLVPPRPPLSORNLQPSRSPILRQDDPVFRTVSPT 202
 OY 181 CSDTAGSCSDSGTSLRPAVPAISILTSNOSSLAPVSGNPSAPAPVDTITATVISIV 240
 DB 203 CSDTAGSCSDSGTSLRPAVPAISILTSNOSSLAPVSGNPSAPAPVDTITATVISIV 244
 OY 241 HSEKSDGPPROSDNKLPAALTLSEHSKASIAITALEENG----VPGTSPMNSIPVG 296
 DB 245 HSEKSDGPPROSDNKLPAALTLSEHSKASIAITALEENG----VPGTSPMNSIPVG 300
 OY 297 QLLHGOQLPTTGPVIRINARPGGPRGHEHNSKLGVLVSPQOILRGSRCSYRSINRGR 356
 DB 301 QLLHGOQLPTTGPVIRINARPGGPRGHEHNSKLGVLVSPQOILRGSRCSYRSINRGR 360
 OY 357 HHSERSORTQSPSLPAPCPVPPRPPLVPPPHLPLRPGVPPQSPQPPSSOPPTAG 416
 DB 361 HHSERSORTQSPSLPAPCPVPPRPPLVPPPHLPLRPGVPPQSPQPPSSOPPTAG 420
 OY 417 YSVPPGPPAPANISTACSPGCVPLAHNTMPTDAPLISREEFREQ-----465
 DB 421 YSVPPGPPAPANISTACSPGCVPLAHNTMPTDAPLISREEFREQ-----480
 OY 466 NDKGRE-----SKPPYSGSSYSRSSTYDS-SQGLAQHIAHLT 501
 DB 481 KLDEFNTDEKELMEYKKIOKERRRFSNSKSYSSSSYSRSSTYDS-SQGLAQHIAHLT 540
 OY 502 LSPSAHTLDLHDHPH-----PEEAARSAMIVHMPDLMDIAHARSRSPPRYRSR 557
 DB 541 RSPSRSHRSRYSRSPRYPRGRGRKSNRYRSRSH-----GYHRSRSRSPRYRSR 595
 OY 558 SPPEFRGOSPTKRNVPREEKEREYFNRYRPPVPPYDILKATYGSVDFRDFEKERTREME 617

DB 596 SPOAFRGOSPNKRNVPQGETEREYFNRYRPPVPPYDILKATYGSVDFRDFEKERTREME 655
 OY 618 RRYREMYEYKGYAGAPRPSANREDSPERLLPLNTRNSPFTGRREDYAAAGOSHNRN 677
 DB 656 RRYREMYEYKGYAGAPRPSANREDSPERLLPLNTRNSPFTGRREDYAAAGOSHNRN 715
 OY 678 RNLGNVPEKLTSTRDSHNAKNKPNKSKESSENVPGDGKGNKRNKRNNEKEGESESEF 737
 DB 716 RNLGNVPEKLTSTRDSHNAKNKPNKSKESSENVPGDGKGNKRNKRNNEKEGESESEF 772
 OY 738 LNPBLETSSKRCGSSGIDETKTDTLPVLPSPRDATPVDEPMDASITFKSVSDOKRE 797
 DB 773 LNPBLETSSKRCGSSGIDETKTDTLPVLPSPRDATPVDEPMDASITFKSVSDOKRE 832
 OY 798 KOKPKYKSDTKRKSOGSATAKKNVLPKSKGPOEKVDGKERKSPREPLKAKKEATK 857
 DB 833 KOKPKYKSDTKRKSOGSATAKKNVLPKSKGPOEKVDGKERKSPREPLKAKKEATK 892
 OY 858 IDSVKPSSSQKDEKVTGPRKAKSKAKDTRROSOPTRRSRKRYPKTSSQKQVYRTR 917
 DB 893 IDSVKPSSSQKDEKVTGPRKAKSKAKDTRROSOPTRRSRKRYPKTSSQKQVYRTR 950
 OY 918 RPRSLRKINYLAREKN-----EREKR---KSVKDESSSKISKVEGETEIVKPSR 967
 DB 951 RPRSLRKINYLAREKN-----EREKR---KSVKDESSSKISKVEGETEIVKPSR 1004
 OY 968 KRMKEGVEKLETPPEKDKIASSITPAKKIKLNENCKIKGNMASTTEPSEKLEST 1027
 DB 1005 KRMKEGVEKLETPPEKDKIASSITPAKKIKLNENCKIKGNMASTTEPSEKLEST 1063
 OY 1028 SKIOEKYKGRKRVKVGSESSSTLVDTYSTSTGSGPVAKSEKTDTRKTVIKIMEY 1087
 DB 1064 SKIOEKYKGRKRVKVGSESSSTLVDTYSTSTGSGPVAKSEKTDTRKTVIKIMEY 1123
 OY 1088 NNDTAPAEVITIMIQVPOSKWMDKDFESEBEDYKTTPIOSVCKPSSIIINVTTKSAT 1147
 DB 1124 NNDTAPAEVITIMIQVPOSKWMDKDFESEBEDYKTTPIOSVCKPSSIIINVTTKSAT 1183
 OY 1148 AKYTEKESQPEKLOKLPKESHELMQHELRSSGSASEGKAKDRHSGSEKDNPKR 1207
 DB 1184 AKYTEKESQPEKLOKLPKESHELMQHELRSSGSASEGKAKDRHSGSEKDNPKR 1243
 OY 1208 KSGAPDKESTYVDRSEOGHFKTSSQSKETRTSEKHSYVGSNKDTPGRDKKVDYS 1267
 DB 1244 KSGAPDKESTYVDRSEOGHFKTSSQSKETRTSEKHSYVGSNKDTPGRDKKVDYS 1303
 OY 1268 RDYSSSKRDERGELARRKDSPPRKGESLSGOKSLREERDLPKKG-AESKSNSSPPRD 1326
 DB 1304 RDYSSSKRDERGELARRKDSPPRKGESLSGOKSLREERDLPKKG-AESKSNSSPPRD 1363
 OY 1327 KRPDHKAPYEKTRCEETKPPVDKNSGKEREKHAAREARKESG 1371
 DB 1364 KRPDHKAPYEKTRCEETKPPVDKNSGKEREKHAAREARKESG 1408

RESULT 3
 ID 015290 PRELIMINARY: PRT: 948 AA.
 AC 015290:
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE RB PROTEIN BINDING PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-96129310; PubMed-8595913;
 Sakai Y., Saijo M., Coelho K., Kishino T., Niihawa N., Taya Y.;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2002, 16:19:12 ; Search time 69.71 Seconds
(without alignments)
3484.219 Million cell updates/sec

Title: US-09-811-045a-1
Perfect score: 7374
Sequence: 1 MAEYKDPNKKAMLTFTGKY.....WRRSWLLGRWRAPSSRNPS 1404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|----------------------------|
| 1 | 6054 | 82.1 | 1591 | 11 | P97868 mus musculu |
| 2 | 5451.5 | 73.9 | 1616 | 4 | Q96PH3 Q96ph3 homo sapien |
| 3 | 3500.5 | 47.5 | 948 | 4 | Q15290 Q15290 homo sapien |
| 4 | 1702 | 23.1 | 628 | 4 | Q9H5M5 Q9H5m5 homo sapien |
| 5 | 1305.5 | 17.7 | 529 | 4 | Q9NPK4 Q9npk4 homo sapien |
| 6 | 542 | 7.4 | 1231 | 5 | Q9XZ21 Q9xz21 drosophila |
| 7 | 480.5 | 6.5 | 1190 | 5 | Q6Z235 Q6z235 caenorhabdi |
| 8 | 405 | 5.5 | 2081 | 10 | Q91H98 Q91h98 arabidopsis |
| 9 | 396.5 | 5.4 | 2752 | 4 | Q9U035 Q9u035 homo sapien |
| 10 | 396 | 5.4 | 4880 | 11 | Q9JL71 Q9jlt1 rattus norv |
| 11 | 388.5 | 5.3 | 5120 | 13 | Q9PU36 Q9pu36 gallus galli |
| 12 | 384 | 5.2 | 3484 | 5 | P91257 P91257 caenorhabdi |
| 13 | 377.5 | 5.1 | 2296 | 4 | Q9UHA8 Q9uhab8 homo sapien |
| 14 | 375 | 5.1 | 17352 | 5 | Q95YM2 Q95ym2 procambarus |
| 15 | 374.5 | 5.1 | 5327 | 5 | Q76891 Q76891 drosophila |

| | | | | | |
|----|-------|-----|------|----|---------------------------|
| 17 | 367.5 | 5.0 | 4833 | 11 | Q9QYX6 Q9qyx6 mus musculu |
| 18 | 367.5 | 5.0 | 5038 | 11 | Q9QYX7 Q9qyx7 mus musculu |
| 19 | 363.5 | 4.9 | 6632 | 5 | Q17362 Q17362 caenorhabdi |
| 20 | 363.5 | 4.9 | 6642 | 5 | Q01761 Q01761 caenorhabdi |
| 21 | 360.5 | 4.9 | 2083 | 5 | Q9N435 Q9n435 caenorhabdi |
| 22 | 351 | 4.8 | 2768 | 5 | Q9VC00 Q9vc00 drosophila |
| 23 | 349 | 4.7 | 990 | 13 | Q91803 Q91803 xenopus lae |
| 24 | 344.5 | 4.7 | 3111 | 5 | Q9VH10 Q9vh10 drosophila |
| 25 | 341.5 | 4.6 | 1422 | 6 | Q95KM4 Q95km4 canis famli |
| 26 | 340 | 4.6 | 2703 | 5 | Q9VPE9 Q9vpe9 drosophila |
| 27 | 340 | 4.6 | 5322 | 5 | Q9N164 Q9n164 drosophila |
| 28 | 339 | 4.6 | 2722 | 5 | Q19135 Q19135 caenorhabdi |
| 29 | 338.5 | 4.6 | 2157 | 11 | Q9ZIR1 Q9zir1 mus musculu |
| 30 | 336.5 | 4.6 | 4498 | 13 | Q93291 Q93291 fuqu rubrip |
| 31 | 336 | 4.6 | 1062 | 5 | Q960C4 Q960c4 drosophila |
| 32 | 335.5 | 4.5 | 933 | 4 | Q9UQ14 Q9uq14 homo sapien |
| 33 | 335.5 | 4.5 | 1026 | 4 | Q9UJ57 Q9uj57 homo sapien |
| 34 | 333 | 4.5 | 1108 | 5 | Q9ND10 Q9nd10 babesia big |
| 35 | 333 | 4.5 | 1262 | 4 | Q9UQ40 Q9uq40 homo sapien |
| 36 | 333 | 4.5 | 1343 | 4 | Q9H7N4 Q9h7n4 homo sapien |
| 37 | 332.5 | 4.5 | 1151 | 5 | Q9VAV4 Q9vav4 drosophila |
| 38 | 331 | 4.5 | 2897 | 5 | Q9VID9 Q9vid9 drosophila |
| 39 | 331 | 4.5 | 2951 | 5 | Q9W3Z0 Q9w3z0 drosophila |
| 40 | 330 | 4.5 | 1430 | 5 | Q9W0B0 Q9w0b0 drosophila |
| 41 | 329 | 4.5 | 1160 | 11 | Q9ESC8 Q9esc8 mus musculu |
| 42 | 328 | 4.4 | 2701 | 4 | Q9Y520 Q9y520 homo sapien |
| 43 | 327 | 4.4 | 2187 | 11 | P70670 P70670 mus musculu |
| 44 | 326.5 | 4.4 | 1341 | 4 | Q9UKV3 Q9ukv3 homo sapien |
| 45 | 325.5 | 4.4 | 1178 | 5 | Q9NHX5 Q9nhx5 drosophila |

ALIGNMENTS

RESULT 1
P97868 PRELIMINARY; PRT; 1591 AA.
ID AC P97868; P70287;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DR 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RETINOBLASTOMA BINDING PROTEIN 6 (PROLIFERATION POTENTIAL-RELATED PROTEIN) (P53-ASSOCIATED CELLULAR PROTEIN) (PACT).
GN RBBP6 OR P2P-R.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY (ISOFORM 2).
RC STRAIN-BALB/C; PubMed=9037032;
RX MEDLINE=97188447; PubMed=9037032;
RA Witte M.M., Scott R.E.;
RT "The proliferation potential protein-related (P2P-R) gene with domains encoding heterogeneous nuclear ribonucleoprotein association and Rb1 binding shows repressed expression during terminal differentiation.";
RT Proc. Natl. Acad. Sci. U.S.A. 94:1212-1217(1997).
RL [2]
RN REVISIONS.
RP STRAIN-BALB/C;
RC Witte M.M., Scott R.E.;
RX MEDLINE=97163396; PubMed=9010216;
RA Simons A., Melamed-Bessudo C., Wolkowicz R., Sperling J., Sperling R., Eisenbach L., Rotter V.;
RT "PACT: cloning and characterization of a cellular p53 binding protein that interacts with Rb.";
RT Oncogene 14:145-155(1997).
CC -!- FUNCTION: BINDS TO Rb1 THROUGH THE Rb1 POCKET DOMAIN. ASSOCIATES WITH HNRNP. ALSO BINDS TO P53. MAY BE INVOLVED IN GROWTH CONTROL

```
CC AND DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS; ISOFORM 1 (SHOWN HERE) AND
CC ISOFORM 2: MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN TESTIS.
CC MODERATE LEVELS FOUND IN HEART, LUNG, LIVER AND SKELETAL MUSCLE.
CC VERY LOW LEVELS DETECTED IN KIDNEY, BRAIN AND SPLEEN. ALSO
CC EXPRESSED IN GROWING UNDIFFERENTIATED MESENCHYMAL STEM CELLS BUT
CC NOT DURING THE TERMINAL DIFFERENTIATION PHASE.
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR EMBL: U03913; AAC72432.1; -.
DR EMBL: U28789; AAB49620.1; -.
DR MGI: MGI:894835; Rbdp6.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: PF00097; zf-C3HC4; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS00518; ZINC_FINGER_C3HC4; FALSE_NEG.
KW Zinc-finger; Metal-binding; Nuclear protein;
KW Alternative splicing.
FT ZN_FING 61 101 C3HC4-TYPE.
FT DOMAIN 481 540 ARG/SER-RICH.
FT VARSLIC 454 487 LYS-RICH (BASIC).
FT CONFLICT 35 35 MISSING (IN ISOFORM 2).
FT CONFLICT 118 119 F -> I (IN REF. 1).
FT CONFLICT 142 142 RQ -> GR (IN REF. 1).
FT CONFLICT 219 219 H -> P (IN REF. 1).
FT CONFLICT 222 222 S -> F (IN REF. 1).
FT CONFLICT 381 381 S -> V (IN REF. 1).
FT CONFLICT 381 381 P -> L (IN REF. 1).
FT CONFLICT 396 396 P -> T (IN REF. 1).
FT CONFLICT 416 423 PWSGVO -> ACSPGVP (IN REF. 1).
FT CONFLICT 430 430 I -> M (IN REF. 1).
FT CONFLICT 437 437 P -> L (IN REF. 1).
FT CONFLICT 448 448 R -> K (IN REF. 1).
FT CONFLICT 490 490 S -> F (IN REF. 1).
FT CONFLICT 504 504 Y -> D (IN REF. 1).
FT CONFLICT 590 590 Q -> R (IN REF. 1).
FT CONFLICT 742 742 R -> RNEE (IN REF. 1).
FT CONFLICT 757 759 ETS -> GKE (IN REF. 1).
FT CONFLICT 764 764 E -> G (IN REF. 1).
FT CONFLICT 779 779 L -> F (IN REF. 1).
FT CONFLICT 783 783 D -> E (IN REF. 1).
FT CONFLICT 813 813 N -> K (IN REF. 1).
FT CONFLICT 1091 1091 K -> T (IN REF. 1).
FT CONFLICT 1117 1117 Q -> H (IN REF. 1).
FT CONFLICT 1365 1365 N -> I (IN REF. 1).
FT CONFLICT 1382 1382 F -> D (IN REF. 1).
FT CONFLICT 1392 1392 P -> L (IN REF. 1).
FT CONFLICT 1397 1397 P -> L (IN REF. 1).
FT CONFLICT 1405 1405 A -> V (IN REF. 1).
SQ SEQUENCE 1591 AA; 177763 MW; 7D0C6798F49D3748 CRC64;
```

```
Query Match 82.1%; Score 6054; DB 11; Length 1591;
Best Local Similarity 84.7%; Pred. No. 1.3e-309;
Matches 1199; Conservative 31; Mismatches 111; Indels 74; Gaps 10;
```

```
QY 1 MMEVKDPMNMGALMTNTGKAIPTIDAFAAIGKKEKPPLEPSSSSSEDDPIPAELL 60
DB 1 MMEVKDPMNMGALMTNTGKAIPTIDAFAAIGKKEKPPLEPSSSSSEDDPIPAELL 60
DB 61 CLICKIMTDAVVIPEGNSGDECIPTALLESDEHTCPTCHONDVSPDALIANKFLROA 120
QY 61 CLICKIMTDAVVIPEGNSGDECIPTALLESDEHTCPTCHONDVSPDALIANKFLROA 120
DB 61 CLICKIMTDAVVIPEGNSGDECIPTALLESDEHTCPTCHONDVSPDALIANKFLROA 120
QY 121 VNNFKMETGTRKRLRQLPFLVPPRRPLSQNRNQPRRSRPLLRQODVVRVRYVSPT 180
DB 121 VNNFKMETGTRKRLRQLPFLVPPRRPLSQNRNQPRRSRPLLRQODVVRVRYVSPT 180
QY 121 VNNFKMETGTRKRLRQLPFLVPPRRPLSQNRNQPRRSRPLLRQODVVRVRYVSPT 180
DB 121 VNNFKMETGTRKRLRQLPFLVPPRRPLSQNRNQPRRSRPLLRQODVVRVRYVSPT 180
QY 181 CSPTKTAGSCSDSGTL SRLPAPSISLTSMQSLAPVSGNSAPAPVDTATVSIYSV 240
DB 181 CSPTKTAGSCSDSGTL SRLPAPSISLTSMQSLAPVSGNSAPAPVDTATVSIYSV 240
DB 181 HS-----APSIISLTSMNPALAPSVSGNPSAPAPVDTATVSIYS 222
```

```
QY 241 HSEKSDGPFRRDNDKLLPAALALSHSKGASIAITALEMEKG----VPGTSPWNSIFVG 296
DB 223 HSEKSDGPFRRDNDKLLPAALALSHSKGASIAITALEMEKGQYVPLGTPS----LLG 278
QY 297 QSLIHGOLITTPGVRIINARPGGRRGWEHSHKGLVYSPPOQIRGENSCSTINRGR 356
DB 279 QSLIHGOLITTPGVRIINARPGGRRGWEHSHKGLVYSPPOQIRGENSCSTINRGR 358
QY 357 HHSERSORTSPSLATPCVPPPPPLVPPPHTLPLPGVPPPOFSPPOFSPPTGAP 416
DB 339 HHSERSORTSPSLATPCVPPPPPLVPPPHTLPLPGVPPPOFSPPOFSPPTGAP 416
QY 417 YSVPPGFPFAPANISTACSPGVPANSTMDPTOAPLLSREEFYREQ----- 465
DB 399 YSVPPGFPFAPANISTACSPGVPANSTMDPTOAPLLSREEFYREQ----- 465
QY 466 -----NDKRE-----SKFYSSGSSYSSSYTS-SQGLAQHIAL 501
DB 459 KLEDFNDFAKELMEYKKIQKERRRSPFSKSPYSSGSSYSSSYTSKSSGSTRSYS 518
QY 502 LSPSAHTLDLHDHPH-----PEEAARSGAMIVHMDLMDIAHARSPPRYRRSR 557
DB 519 RFSRSHSRYSRSPYPRRGGRKSRNYSRSH-----GYHRSRSPRYRRSR 573
QY 558 SPPEFPGQSPTRKRVNPREKEREYFNRYREPPYDIKAYGSRVDFDPFEKERYREWE 617
DB 574 SPQAFRGQSPTRKRVNPREKEREYFNRYREPPYDIKAYGSRVDFDPFEKERYREWE 633
QY 618 RKTREYETKRYGVAQRPANREDSPERLLPLNTRNSFTGRRDYAAGSHRN 677
DB 634 RKTREYETKRYGVAQRPANREDSPERLLPLNTRNSFTGRRDYAAGSHRN 693
QY 678 RNIGNVPEKLTSRDISHNAKNPKSEKESENVPGDGKGNKHKHRRRNEEKGSESE 737
DB 694 RNIGNVPEKLTSRDISHNAKNPKSEKESENVPGDGKGNKHKHRRRNEEKGSESE 750
QY 738 LNPelletSKRCGSSGIDETKTDTLFLVPSRDATPVREDPDAISTFYKVSXDKRE 797
DB 751 LNPelletSKRCGSSGIDETKTDTLFLVPSRDATPVREDPDAISTFYKVSXDKRE 810
QY 798 KDKPKYKSDTKRKSOGSAATKADNVLPKSPQOEYVDGREGSPSEPLKAKKEATK 857
DB 811 KDKPKYKSDTKRKSOGSAATKADNVLPKSPQOEYVDGREGSPSEPLKAKKEATK 870
QY 858 IDSVKSSSSQKDEKTYGPRKAKSHAKDTRROSQRTRRSKRYVKTSSQSOQVTR 917
DB 871 IDSVKSSSSQKDEKTYGPRKAKSHAKDTRROSQRTRRSKRYVKTSSQSOQVTR 930
QY 918 RPRSLKINTLIR-EKNERKKKSYVDKDFESSMKISVETETLYKPSPKKMEGDVE 976
DB 931 KAKPRKKNKLDKSGKRRKRTKEESVDKDFESSMKISVETETLYKPSPKKMEGDVE 990
QY 977 KLERTEPKDIASSTPAKKIKLNRETGKKIGNAENASTTKEPESEKLESTSSKIKOEYK 1036
DB 991 KLERTEPKDIASSTPAKKIKLNRETGKKIGNAENASTTKEPESEKLESTSSKIKOEYK 1050
QY 1037 GKAKRRVAGSESSSTLVDTYSTSGSPVRKSEKTDTRKTVIKTMEBYNDNTAPAE 1096
DB 1051 GKAKRRVAGSESSSTLVDTYSTSGSPVRKSEKTDTRKTVIKTMEBYNDNTAPAE 1110
QY 1097 DVILMIQVQSKNDKDFESEBEDVKTTPIQGVGRPSSTIKNVTKRPSATAYTKESE 1156
DB 1111 DVILMIQVQSKNDKDFESEBEDVKTTPIQGVGRPSSTIKNVTKRPSATAYTKESE 1170
QY 1157 QPEKLOKLPKEASHLMOHELRSKSGASSEKRAKDREHSGSEKNDPKRKSGAOPDE 1216
DB 1171 QPEKLOKLPKEASHLMOHELRSKSGASSEKRAKDREHSGSEKNDPKRKSGAOPDE 1230
QY 1217 STVDRLSQEGHFKTLLSQSSKERTSEKHESVRSNNKDFTPGRKRVYDSDRYSKSKR 1276
DB 1231 STVDRLSQEGHFKTLLSQSSKERTSEKHESVRSNNKDFTPGRKRVYDSDRYSKSKR 1290
QY 1277 DERGELARRKDSPPRGKESLSGQSKLREBRDLPKRGAESKKSNSPPRDKKHHKAP 1336
```

Db 1291 DERGELARRKDSPPRGKESLSQSKLEBERDLPKKGAESKSNSSPPRDKKPHDHKAVY 1350
QY 1337 ETKRCEETKPPVDKNSGKERKHAABARNGKSSG 1371
Db 1351 ETKRCEETKPPVDKNSGKERKHAABARNGKSSG 1385
RESULT 2
ID Q96PH3 PRELIMINARY; PRT; 1616 AA.
AC Q96PH3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROLIFERATION POTENTIAL-RELATED PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Camargo A.A., Moreira E.S., Simpson A.J.G.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF352051; AAL05625.1; -
SQ SEQUENCE 1616 AA; 182248 MW; D89F5BA8ED95D72B CRC64;

Query Match 73.9%; Score 5451.5; DB 4; Length 1616;
Best Local Similarity 75.3%; Pred. No. 5,6e-278;
Matches 1073; Conservative 92; Mismatches 167; Indels 93; Gaps 16;

QY 1 MNEVDPNNKGMALNTGKYAIPITDAEYATGKKEKPPFLDEBSSSEEDDPPEALL 60
Db 23 MNEVDPNNKGMALNTGKYAIPITDAEYATGKKEKPPFLDEBSSSEEDDPPEALL 82
QY 61 CLICDIMTDAVYIPCCGSSCDECIRTLLESDEKHTCTCHONDVSPALANKFLROA 120
Db 83 CLICDIMTDAVYIPCCGSSCDECIRTLLESDEKHTCTCHONDVSPALANKFLROA 142
QY 121 VNNFKNETGYTRLRKQLPFLFLVPPPPPLSQRLQPPRSPTILNQDPVVFRTVSPT 180
Db 143 VNNFKNETGYTRLRKQLPFLFLVPPPPPLSQRLQPPRSPTILNQDPVVFRTVSPT 202
QY 181 CSDTATAGCSGSLSLRPPAPSSISLTSSQSLAPPVGNSAPAPPPDITATYSIV 240
Db 203 -----HPAPSSISLTSSQSLAPPVGNSAPAPPPDITATYSIV 244
QY 241 HSEKSDGPPRDSNDKLLPAAALTSEHSKASSIATATLMEKG---VGTSPMNSIFVG 296
Db 245 HSEKSDGPPRDSNDKLLPAAALTSEHSKASSIATATLMEKG---VGTSPMNSIFVG 300
QY 297 QSLHGGQILLPTTGVPYRINARPGGGRGWEHSHNKLGLYLVSPQIIRGERSCYRINRG 356
Db 301 QSLHGGQILLPTTGVPYRINARPGGGRGWEHSHNKLGLYLVSPQIIRGERSCYRINRG 360
QY 357 HSEHSQRTQSLSLPTPCGFVPPPPPLPLPPLPPLPPLPPLPPLPPLPPLPPLPPL 416
Db 361 HSEHSQRTQSLSLPTPCGFVPPPPPLPLPPLPPLPPLPPLPPLPPLPPLPPLPPL 420
QY 417 YSVPPGPPAPANISTACFSPGVPFAHSNTMPTTQAPLSREEFREBO-----465
Db 421 YSVPPGPPAPANISTACFSPGVPFAHSNTMPTTQAPLSREEFREBO-----465
QY 466 -----NDKGR-----SKPPYSGSYSRSSYTDSS-SQGLAQHIALVL 501
Db 481 KLDEFNDAKELMEYKTIQKERRRFSRSKSPYSGSSYSRSSTYTSKRSRGSTRSRYS 540
QY 502 LSPSAHTIDLDLHDPH-----PEEAARSAMVHMPDLMDTAHAASRSPRYRYSR 557
Db 541 RSPFSRSHSYSRSPYPRGRGKSRNYSRSH-----GYHRSRSPRYRYSR 595
QY 558 SPPEFGSGPTKRNVPREKEREYFNRYREVPPPYDIKAYGGSVDFRDPFEKERREWE 617

Db 596 SPQAFRGSGPNKRNVPQGETEREYFNRYREVPPPYDMKAYGGSVDFRDPFEKERREWE 655
QY 618 KYREMYEYKYGYAVGAQPPRSANRDEFSPELLPLINRNSPFTRGREDYAAQOSHRN 677
Db 656 KYREMYEYKYGYAVGAQPPRSANRDEFSPELLPLINRNSPFTRGREDYAAQOSHRN 715
QY 678 RNLGGNYPEKLTSPRSHANKDNPKSEKSEENVPGDGKGNKKHKKRRNEEKGEESEF 737
Db 716 RNLGGNYPEKLTSPRSHANKDNPKSEKSEENVPGDGKGNKKHKKRRNEEKGEESEF 772
QY 738 LNPelletSRKCRSGSGLDEKTDLTFLVPSRSDATPVDEPDAESTFKSVSDKRE 797
Db 773 LNPelletSRKCRSGSGLDEKTDLTFLVPSRSDATPVDEPDAESTFKSVSDKRE 832
QY 798 KKKPKYKDKTRKSDGATAKDNVLAASKGPOEKVDGDRKSPRSEPLKAKAEATK 857
Db 833 RUKPRAKGDTRKRNKNSVSKKENYVPAKGPQEKVDGERRSPRSEPLKAKAEATK 892
QY 858 IDSVPRSSSOKDEKVTGTPRAKHSKADTRRQSPRTRSKRTVPKTSQKSOQVPR 917
Db 893 TDTKSSSSSOKDEKVTGTPRAKHSKADTRRQSPRTRSKRTVPKTSQKSOQVPR 950
QY 918 RPSRLKINYLILAREKN-----ERERK---KSYDKDFESSSMKISVETETIYKPS 967
Db 951 EKA-KRP-----EKNKPLDMKGEKRRKTEKGYDKDFESSSMKISVETETIYKPS 1004
QY 968 KRMGDEVEKLETPBKDKIASSTTPPAKIKILNREGKIKGNAENASTKEPSEKLETS 1027
Db 1005 KRMGDEVEKLETPBKDKIASSTTPPAKIKILNREGKIKGNAENASTKEPSEKLETS 1063
QY 1028 SKIKOEKVGKAKRKVASEGSSSTLVDTSTSTGSGVVRKSEKTDKRVYKMEY 1087
Db 1064 SKIKOEKVGKAKRKVASEGSSSTLVDTSTSTGSGVVRKSEKTDKRVYKMEY 1123
QY 1088 NNDNTAPADVYIIMIQVPSKMDKDFESEEDVKTQPIQSGKSSIIKNTYTPSAT 1147
Db 1124 NNDNTAPADVYIIMIQVPSKMDKDFESEEDVKTQPIQSGKSSIIKNTYTPSAT 1183
QY 1148 AKYTEKESQPEKLOKLPKASHLMOHELRSKSGASSEKGRADREHSGSEKNDPKR 1207
Db 1184 AKYTEKESQPEKLOKLPKASHLMOHELRSKSGASSEKGRADREHSGSEKNDPKR 1243
QY 1208 KSGAOPDKESTYDRLSEOGHFTLSOSKETRSEKHEBVSRSNKPDPFGDKYDVDS 1267
Db 1244 KSTQPEKESNLDRLNEOGNEFSLSSKEARTSDKHDSSTRASNKDPFNKDKTDYPT 1303
QY 1268 RYSSSKRRDERGELARRKDSPPRGKESLSQSKLEBERDLPKKG-AESKKNSSPPRD 1326
Db 1304 RYSSSKRRDERGELARRKDSPPRGKESLSQSKLEBERDLPKKG-AESKKNSSPPRD 1363
QY 1327 KRPDHKADYETKRCCEETKPPVDKNSGKERKHAABARNGKSSG 1371
Db 1364 KRPDHKADYETKRCCEETKPPVDKNSGKERKHAABARNGKSSG 1408
RESULT 3
ID Q15290 PRELIMINARY; PRT; 948 AA.
AC Q15290;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RB PROTEIN BINDING PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-96129310; PubMed-8595913;
Sakai Y., Saijo M., Coelho K., Kishino T., Nakawa N., Taya Y.;

RT "cDNA sequence and chromosomal localization of a novel human protein,
RT RBQ-1 (RBQB6), that binds to the retinoblastoma gene product.";
RT Genomics 30:98-101(1995).
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL: X85133; CAAS9445.1;
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF00097; zfc-C3HC4.1.
DR SMART: SM00184; RING; 1.
DR ZINC-finger.
KW SEQUENCE 948 AA; 107156 MW; 5F24EBD5E9E340A2 CRC64;
SQ
Query Match 47.5%; Score 3500.5; DB 4; Length 948;
Best Local Similarity 73.2%; Pred. No. 7.2e-176;
Matches 691; Conservative 44; Mismatches 130; Indels 79; Gaps 10;
QY 1 MMEVKDPNNKGMALITNGKYYAIPITDAEAYVAGKKEKPPFLPEPSSSSSEDDPIAEILL 60
DB 23 MMEVKDPNNKGMALITNGKYYAIPITDAEAYVAGKKEKPPFLPEPSSSSSEDDPIAEILL 82
QY 61 CLICKIMTDVAVIIPCCGSSCDECIPTILLESDEKTCPTCHQNDVSPALLANKFLRQA 120
DB 83 CLICKIMTDVAVIIPCCGSSCDECIPTILLESDEKTCPTCHQNDVSPALLANKFLRQA 142
QY 121 VNNFKETGTYTRKQOLPEFLVPPRPLSQNLQPSRSRSPILRQDPVFRVTVSPT 180
DB 143 VNNFKETGTYTRKQOLPEFLVPPRPLSQNLQPSRSRSPILRQDPVFRVTVSPT 202
QY 181 CSDPTKTAGSCSDSGLSLRLPAPSISSILTSNOSSIAPVSGNPSADAPVDITATVISIV 240
DB 203 CSDPTKTAGSCSDSGLSLRLPAPSISSILTSNOSSIAPVSGNPSADAPVDITATVISIV 244
QY 241 HSEKSGPPRDSNKLIPAAALTSEHSGKSSAITATLAMEKG----VPGTSMWNSIPVG 296
DB 245 HSEKSGPPRDSNKLIPAAALTSEHSGKSSAITATLAMEKG----VPGTSMWNSIPVG 300
QY 297 QSLHLQOLITPTGPVRINARPGGPRGMEHSHKGLGYLSPPOIRGERSCYRSINGR 356
DB 301 QSLHLQOLITPTGPVRINARPGGPRGMEHSHKGLGYLSPPOIRGERSCYRSINGR 360
QY 357 HHSERSQRTQSPSLPATPCFVYVPPPLYPPLHTLPLPGVPPRQSPQFSSQPTAG 416
DB 361 HHSERSQRTQSPSLPATPCFVYVPPPLYPPLHTLPLPGVPPRQSPQFSSQPTAG 420
QY 417 YVPPRPPRPPANISACSPGVPATSHNTMPTOAPLISREEFYEQ-----465
DB 421 YVPPRPPRPPANISACSPGVPATSHNTMPTOAPLISREEFYEQ-----460
QY 466 -----NDKGR-----SKFYSGSSSYRSSTYDS--SGLAQHIALT 501
DB 481 KIDEFTNDFAKELMEYKKIKERRRSFSKSKSPYSGSSSYRSSTYDS--SGLAQHIALT 540
QY 502 LSPSAHTLDDLHDHPH---PEEAARSAMVHMPDLMDIAHARSPPRYRYSR 557
DB 541 RFSRSRSHRSYSRSPYPRGRGSKSRNYSRSH---GYHRSRSPRYRYSR 595
QY 558 SPPEFRQSTTKRNVPREREKRETFNRYRVPYDIAKAYGSRVDRDPEKRYEME 617
DB 596 SPPEFRQSTTKRNVPREREKRETFNRYRVPYDIAKAYGSRVDRDPEKRYEME 655
QY 618 RKYREWEKYYKGYAVAGORPSANREDSPERLLPLINRSPTRGRRREDYAGAGSHN 677
DB 656 RKYREWEKYYKGYAVAGORPSANREDSPERLLPLINRSPTRGRRREDYAGAGSHN 715
QY 678 RNLGANTPEKLRSDSHNAKDNPKSEKESBENYPGDGKGNKHKRRRNEKEGESESF 737
DB 716 RNLGANTPEKLRSDSHNAKDNPKSEKESBENYPGDGKGNKHKRRRNEKEGESESF 772
QY 738 LNPBLETSRKRGSSGIDETKTDTLFLVDSRDDATPVDRDPMDESIITTKSVADKRE 797
DB 773 LNPBLETSRKRGSSGIDETKTDTLFLVDSRDDATPVDRDPMDESIITTKSVADKRE 832
QY 798 KDKFKVSDKTKRKSDDSAKAKDNVLPKSGQEKVDGDEKPRSEPPPLKAKKEATK 857

DB 833 RDKPRAKGDGKTRKNDGSAVSKKENIVKPAKGPQEKVGDVBDLIDLNLQKKPRRLR 892
QY 856 IDSVPRSSSQDEKVTGTPRRRAHKSADKTRQSQPRTSRSKR 901
DB 893 LITLNMHLPLRRMKSLPEPEKL-----TLNQOKTPRKNITSOR 930
RESULT 4
QY 09H5M5 PRELIMINARY; PRT; 628 AA.
AC 09H5M5;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CDNA: FLJ23301 FIS, CLONE HEP1120.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hiki J. T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Oca T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RA "NEO human cDNA sequencing project."
RT Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AK026954; BAB15600.1;
SQ SEQUENCE 628 AA; 70907 MW; F9FD5998518CC9A CRC64;
Query Match 23.1%; Score 1702; DB 4; Length 628;
Best Local Similarity 78.6%; Pred. No. 1.1e-81;
Matches 331; Conservative 40; Mismatches 48; Indels 2; Gaps 2;
QY 952 MKTSYEGEIVKPSRRKMEGVKLEPTEPKDITASTTPPAKTKILNRETKKIGNAE 1011
DB 1 MKTSYEGEIVKPSRRKMEGVKLEPTEPKDITASTTPPAKTKILNRETKKIGNAE 59
QY 1012 NASTKEPSEKLESTSSKIKOEKVKAKRKVAGSESSSTLVDTYSTSGSPYRKSE 1071
DB 60 NISNTEPSEKLESTSSKIKOEKVKAKRKVAGSESSSTLVDTYSTSGSPYRKSE 119
QY 1072 EKTDTKRTVITKIMEYVNDNTAPAEVYITMIOYPOSKMKDDPESEEDVYKTOPQSVG 1131
DB 120 EKTDTKRTVITKIMEYVNDNTAPAEVYITMIOYPOSKMKDDPESEEDVYKTOPQSVG 179
QY 1132 KPSSITKNTTKSATPAKYTEKESQEPKLOKLPEKASHELMQLHLSRSGSASEKGRA 1191
DB 180 KPASVILKINVTSPKSNIVKYPEKESQEPKLOKLPEKASHELMQLHLSRSGSASEKGRA 1239
QY 1192 KOREHSGSEKDNPDKRRSGAQPDKESTVDRLSEQGHFKTLSSQSKETRITSEKHEVAGSS 1251
DB 240 KDRDYSLLEKNEBKRKNSIQPEKESLMDLNLQGNKSLSSQSKKARTSDKHSTRASS 299
QY 1252 NKDFTPGDRKVDYDSHDYSSSRKRDGELARRKDSPPRGKESLSQSKKLAEBRDLPK 1311
DB 300 NKDFTPGDRKVDYDSHDYSSSRKRDGELARRKDSPPRGKESLSQSKKLAEBRDLPK 359
QY 1312 KG-AESKKSNSPPRODKKPHDHAPIYTKRRCETKTVYDKNNSGKERKHAABRANGESS 1370
DB 360 KGTGDSKKSNSPPRODKKPHDHAPIYTKRRCETKTVYDKNNSGKERKHAABRANGESS 419
QY 1371 G 1371
DB 420 G 420
RESULT 5
QY 09NPX4 PRELIMINARY; PRT; 529 AA.
AC 09NPX4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)


```

DR 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DE 01-OCT-2000 (TREMBlrel. 15, last annotation update)
DI HYPOTHEICAL.59.9 KDA PROTEIN (FRAGMENT).
GN DKEP761B2423.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=AMYGDALA;
RA Blocker H., Boecher M., Brandt P., Mewes H.W., Well B., Wiemann S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL359564; CAB94869.1; -.
KM Hypothetical protein.
FT NON_TER
SQ SEQUENCE 529 AA; 59897 MW; 8F5A66CDF74CCA66 CRC64;

Query Match 17.7%; Score 1305.5; DB 4; Length 529;
Best Local Similarity 77.9%; Pred. No. 5,5e-61;
Matches 250; Conservative 32; Mismatches 38; Indels 1; Gaps 1;

OY 1052 TLVDYTSSTGSGSPYKSEKEDTKRYIKTMEENNDNTAPAEVDYITIOVPOSKWPK 1111
DB 1 TLVDYTSSTGSGSPYKSEKEDTKRYIKTMEENNDNTAPAEVDYITIOVPOSKWPK 60
OY 1112 DFESEEDYKTPQIQVSGKPSIIKNTYTPSATATKTEKSEOPKLOKLPKASHE 1171
DB 61 DFESEEDYKTPQIQVSGKPSIIKNTYTPSATATKTEKSEOPKLOKLPKASHE 120
OY 1172 LM0ELRSKSGSASSSEKAKDREHSGSEKNDPKRKSAGOPKSTVRLISQGFKL 1231
DB 121 IIOHEKSSKSNASSSEKGTDRDYSLKEENPEKRNSTQPKESNLRLNQGPKSL 180
OY 1232 SOSSEKTRSEKESVYSGSSNKPDPGPKDYDSDYSSSKRRDREGELARRKDSPPR 1291
DB 181 SOSSEKTRSEKESVYSGSSNKPDPGPKDYDSDYSSSKRRDREGELARRKDSPPR 240
OY 1292 GRESLSGKSKLREERDLPRKG-AESKKSNSPPRDKPHDKAPYETKRCCEETRPVVK 1350
DB 241 NDSASGCKNKPREEERDLPRKGAGSKNSPSRDKPHDKAPYETKRCCEETRPVVK 300
OY 1351 NSGKERKHAAREARNGKSSG 1371
DB 301 NPKCKREKHAAREARNGKSSG 321

RESULT 6
OYXZ21 PRELIMINARY; PRT; 1231 AA.
AC 09XZ21;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE BCDA:ID21643 PROTEIN.
GN BCDA:ID21643 OR CG3231.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Adamaites P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

```

```

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle S., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Garidelian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Jatali M., Kaulush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster." ;
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
RA Agbayani A., Arcalata T.T., Baxter E., Blazey R.G., Butenhoff C.,
RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Fritse E.,
RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleby J.M.,
RA Park S., Sequeira A., Sethi H., Snit E., Svitskas R.R., Weinburg T.,
RA Celinker S.E.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL: AE003463; AAF47162.1; -.
DR EMBL: AF132177; AAD34765.1; -.
DR Flybase: FBgn027522; BCDA:ID21643.
DR InterPro: IPR001878; Znf_CCHC.
DR InterPro: IPR001841; Znf_CCHC.
DR Pfam: PF00097; zf-C3HC4; 1.
DR Pfam: PF00098; zf-C3HC4; 1.
DR SMART: SM00184; RING; 1.
DR SMART: SM00343; Znf_C2HC; 1.
KW Hypothetical protein; Zinc-finger.
SQ SEQUENCE 1231 AA; 139038 MW; 4A91FF172BDEAB0 CRC64;

Query Match 7.4%; Score 542; DB 5; Length 1231;
Best Local Similarity 21.4%; Pred. No. 1.7e-20;
Matches 301; Conservative 188; Mismatches 434; Indels 486; Gaps 63;

OY 35 KEKPEFLPER-----SSSEEDDPIPAELLCLICKDITWDVAVIPCCGSSSCDCERT 88
DB 188 ROKPDAENSAFVLPVAVONOE---IPEDLLGICGRIDFVDAVMIPCCSSSFCDDCYRT 244
OY 89 TLLESDKHTCPTCHQNDVSPDALIANKFLRQAVNNFNKNETGYTKRLKQLPPLFLVPPP 148
DB 245 SLIESEDECPDCKEKKNKSGSLIPNRLFNVSNAFKNKNGYKNSAK----- 292
OY 149 RPLSQNLQPRSRSPILR--QDDPVYFRYVSPGCDTKAGSCDSGTLSPAPSIS 206
DB 293 -PAAVN--EERPVKEVEKKRVA--EVEEETEVK-----PEKOKES 331
OY 207 ITNSQSLAPPVGNPSSAPAPVDPITATVTSVHSEKSDGPRDSDNKLLPAAALTSEH 266

```

```
Db 332 ETN-----GNNPKSESEPPATTPEPS-QKEKKDYSDYEDNITIKMOPQAPADS-- 379
QY 267 SKGASIAITALEMBEKGVPGTSPMNSIFVGOSLHGOLIPTTGPIVINAARPGGPGPME 326
Db 380 -----TTVPKRS-----PYS 391
QY 327 HSNKGLIVSPPOQIRGECRCYSINGRHHSEKQTOQSPSLPATCFVPPVPLYP 386
Db 392 HRSESSH-----RRDRSDYVDHDKH--ORPSKSESVNKDRS-----427
QY 387 PPPHTLPLPGVPPPGQSPQSPSSOPPTAGYVPPPPGPPAPANISTACFSFGVTAHSN 446
Db 428 -----LPLPPIG-----434
QY 447 TMTPTQABLSEEFYRONDNGREKPEYSGSSYSRSSYTDSSQGLAQHIALTLSPSA 506
Db 435 TLPSYQGHMAESEBAR-----SSAKYKPPYQMOQG-----PPP 469
QY 507 AHTLDLHDHHPPEAARASAMIVHMDLMDIAHAKSRPP--YRKTRSRSPRPPERG 564
Db 470 MH-----MMSHHMPAYNNGFNMGORPLSLYVPYQONOSVHP-----505
QY 565 QSPTRKRVPREKEREYENRYREVPPPYDIKAYYGRSVDFPDPEKERYREMERKYREMY 624
Db 506 -----MRAPYG--SAGGGMNMMSQPFQSPN-----L 530
QY 625 EKYKGYA--VGA-----OPRPSANKEDESPERLLPLNIRNSPFTGRREDYAAGOSHNR 678
Db 531 ASIYGAVAKYGGSPIDDLPLAFNRMKEKRYKVDRRSSDRHRSRPD--RQRHNFK 587
QY 679 NLGNGYR--EKLSTRDSINAKD--NPKSEKESENVPGCGKNKKKH--RKRRENEKGEES 734
Db 588 S-----PAYENDNSD--NLMDKRRSRERKREH-----SYERNRHRRSSRQPPDGSKS 635
QY 735 ESNLPELLETSRRCGSGSIDETKTDLTVLPSRDAATPYRDEPMOAEESTFKSVSKD 794
Db 636 PG-----GRIKR--SG-----HRSASP-----651
QY 795 KREKDKPVKSDKTKRKSDSGSATAKDNVLPKSKG--POEKVGDGREGSPREBPPLKAK 852
Db 652 -----KPGYKSDYDKDPYKNSAPKTEAVEPPPGFELQITD--EDGYNKKHPTJSEA 703
QY 853 EEARKIDSVKSSSSQKDEKVTGPRKAH-----SKSAKDTRRSQOPTRSKKTVER-T 906
Db 704 SOSKSGSSKKRGENRHEE--APRKRHRGRSISKEKPPMDSVNRSLTTPAKITTPKMT 759
QY 907 SSOSQOPVTRRPRSLRT--NYLLAREK-----NEREKKKSVXKDESSSMKISK 956
Db 760 AAOQLRO--RESSPTPEKSHDDYLLAKARIASOPVINDTE--METNNGEKKNAKS--PLSK 815
QY 957 VEGTEIVKPSPRKMEGVEKLEERTPEKDKTASSTTPAKITLNBETGKIKGNAENASTV 1016
Db 816 -----DRKKKKDKDKAER-----KNNKKKRAKKEKGDQKSSS 851
QY 1017 KEPEKELESTSSKTKOEKVKGAARVAGSGSSSTLVYDTSTSTGSGPYRKSEKEDT 1076
Db 852 VNRSDD-----SDINNSLIMNESNYKVLSPRAQSPSI--EINAAQSLSPHNATENNVP 901
QY 1077 KRVYKTMEEVNNNTAPAEVYIIMIQVQSKMDKDESESEEDVKTQPTQSGKPSST 1136
Db 902 KSHSILTVGAASDNLGRSKLSEANSVNLKWEIDENILEO--SSKKAAGASDDPEI 960
QY 1137 IKNTTTPSATATKTEKESEQPERLOKLPRKASHELMOHELRRSSKGSASSEKGRADDEH 1196
Db 961 TSDVLRKAENMIFAKAINALRPMEOVI-----INKKMSKRSV 1000
QY 1197 SGSEKD--NPKRKSQAQPKESTVDRL-----SEQHEKTLSSQSKETRTSEKHESVR 1248
Db 1001 VRSDKDRSSSPRRNNS-----SRSVKDLRTGKISNDSRSKDKSGRRRAARSDDDDANR 1055
QY 1249 GSSKDKTPGDKKVVDSDRYS--SKRDERGELARRKDOPPRGKESLSQKSKLREER 1307
Db 1056 GRSDRHGSRKKDNR-----SRDRAAPSEKROER--STKRSSP-----EDDKLRKON 1099
```

```
QY 1308 DLPKKAESK-----KSNSSPPDDKKPHDHKAPYETK-----RPECETKP-----1347
Db 1100 ---KEQSESKHGKHQDNNSSDSDRRAAKNTKSSDSRVSVYAVAPAPKPCRPDPNPFKE 1156
QY 1348 VDKNSGKER-----EKHAAEARNGKE 1368
Db 1157 VDTSSSSLVVKNYDNTIOKEGASSDNGME 1185

RESULT 7
062235
ID 062235 PRELIMINARY; PRT: 1190 AA.
AC 062235;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F36F2.3 PROTEIN.
CN F36F2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Cottege A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; Pubmed=9851916;
RT none;
RA "genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z81532; CAB04326.2; -.
SQ SEQUENCE 1190 AA; 137239 MW; B219D269962432AB CRC64;

Query Match 6.5%; Score 480.5; DB 5; Length 1190;
Best Local Similarity 19.0%; Pred. No.2.8e-17;
Matches 276; Conservative 168; Mismatches 400; Indels 605; Gaps 51;

QY 1 MME--VKDPNNKGMALNTGKTAIPTI--DAEAYALGKKEKPPPLPEPSSSEDDPIP 56
Db 252 LMETTVDDPD--AMHPSGKYVIPIMHMKARQETLARKNE-----DGSSSPAQTSRKVP 303
QY 57 AELICLICQIDMTAVVIPCQGNSSODECITITLLESDKHTCP--TGHQNVSPDALIAN 114
Db 304 PELLCPICQSLFEKAIYVSCCGNSYCADCEARILDPNOKCPGADCGK-DISTSTIPN 362
QY 115 KFLQAVNNEFNENETGYTKRLKQLPPELFLVPPRPLSQRMIAQRSRSPILRQODPVVFR 174
Db 363 KTLQDAAMAAMLSATGCPGAPPTPOLV-----EPQQR 394
QY 175 YTVSPGSDTYTAGSCSDSGTSLRLPAPSSISLTSNOSSLAPPVSGNBSAPAEVPDITA 234
Db 395 IRIQ-----LKAQS-----SSOSQITP-----411
QY 235 TVTISVHSEKSDGPFRRSDNKLPLPAALITSEHSGASSIATLAMEKGVGTSPMNSIF 294
Db 412 -----SGISPGSTLV 421
QY 295 VGOSLHGOLIPTTGPIVINAARPGGPGMEHNSKGLIVSPPOQIRGERSGYSINR 354
Db 422 QQQFTL-----TSVSS 432
QY 355 GRHHSERSQRTQSPSLPATCFVVP--PPPLYPPTHLLPLPGVP--PQSSPQFPSS 410
Db 433 GTISLAPPSNV--NPSIGIPLASQVPSMVODVSLPPPOLRQELRPGIGLPGF-----484
QY 411 QPTPAGSVPPPGFP-----PAPANISTACSPGVPYTAHSTMTPTQAPLPSRE 459
Db 485 -----GLPPGVPGLSATVLPQHOHSMPLNYGMPLEFSAGPPPAVS-----SVPRPSAISD 533
```


Db 1014 EXEEKSKTKEBAKKEKKKODKREKSE-----RKSKKEKE-- 1053
OY 1115 ESEEDVTTQIQQVGKSSIIKNVTKPSATAYTEKESQPEKIQKLPKEASHE--- 1171
Db 1054 --ESRDLAKK-----KEEETK-----EKKESENHKSKEKEDKHEHEDK 1091
OY 1172 -LMOHELRRSSKGSASSEKGRANDREHSGSEK---DNPKRKSGAOPDKESTYDRLSEQGH 1227
Db 1092 SMKKEEDKKEKKHEESKRKKEEDKDKMEKLEDQNSNKKEDKNEKKKSQ-----H 1143
OY 1228 FRTLSOSKETRTSEKHEVSRRSSNKDFTPGDKKYDYDSROYSSSKRRDEGELARRKD 1287
Db 1144 VKLV--KKESDKKKEKKEEKESEKTEKLEISSQKNEVKKKKSSKDOKKKE----- 1194
OY 1288 SPPRGKESLGGOKSKLREBRDLPKKGAEKSSKSSSPPRDKKPHDHAKAPYETKRPEETKP 1347
Db 1195 -----KEMKESEKKL-----KKNEEDRKOTSYEKKKKOKETKK--EKKNKPKDKKN 1240
OY 1348 VDKNSGKER--EKHAARANGKES 1369
Db 1241 TTKQSGKKKESMESEKKAENOOKS 1265

RESULT 9
OY0Q35 PRELIMINARY; PRT; 2752 AA.
AC OY0Q35:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RNA BINDING PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, complete cds.";
RL Submitted (Jul-1998) to the EMBL/GenBank/DDJr databases.
DR EMBL; AB016092; BAA83718.1; -
DR InterPro; IPR002965; P_tich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 2752 AA; 299672 MW; 109CG64F181097123 CRC64;

Query Match 5.48; Score 396.5; DB 4; Length 2752;
Best Local Similarity 20.28; Pred. No. 1.9e-12;
Matches 362; Conservative 189; Mismatches 661; Indels 577; Gaps 75;

OY 31 AIIKKKEKPLPEEPPSSSEDDPIPAELICLICKIMTDAVVIPOCGNSCDE----- 84
Db 48 ALVKRRPDLIDHEKRR-----VELRCLLEEMEE-----QGYEQQIOEK 90
OY 85 --CIRTTLESKDKHCTCTCHONDVSPDALIANFLKQAV-----NNFKMETGYTKR 133
Db 91 VAFIRLILLE-----KDVNPGKEETPGQRAVAVETHQLAELANKKNE-----R 134
OY 134 LR-----KQLPFLFLVPPRPLLSQRNLQPPRSRSLNKOD 169
Db 135 LRAAFGISDYSVDGSSFPQRRAREAKQAP-----EPPRYSILVRESSSSSTPKOK 189
OY 170 PVFERYTVSPCTSDTKTAGSCSDGT-----LSRLPAPSIS 206
Db 190 K-----KKKHGRKRESSESPRRERKSSKKKKHRSSESKKKRHRSPTRK--- 235
OY 207 LTSNQSILAPVSGNPSSAPAPVDITATVISIVHSEKSDGPFERDSNKLIPAAALTS 266
Db 236 --SKRKSCKDKRKRKRSRSTPAP-----KSRRAHRSTADSASSD-----TSR 276
OY 267 SKGASSIATIALMEKKEVPGTSPWNSTFVGOSLHGQLITTGQVYRINAAARPGGGRPG 326

Db 277 SRSRMAAKT---HTTALAGRSP-----SPASG-----RQEGDAPSE 312
OY 327 HSNKGLYLVSPQOIRRGGRSCYRSTINRGHHHSERQSQSLPATPCFVVPVPPPL-- 384
Db 313 PGTSTORPSSSETATKOPSSPYEDKDKKESATRPSPERSSTGCEPAPPLPLAE 372
OY 385 -----YPPPHPLPL--PCGVPPQFS-----PQFSSQPPAGSYVPPGFPAPANI 431
Db 373 RHGGSFQPLATTPPLSGEPVNPSPSEASPTRDRSPKSPKELPQSSSESSPPS--PQPTKV 430
OY 432 ST-ACFSPGVP-----TAHSNT----- 447
Db 431 SHHASSPESPAPAPAGSHREISSPTSKNRSHGAKRDKSHHTPRMRGRSPATA 490
OY 448 -----MPTQAPLISDEEYREOND-----KGR----- 470
Db 491 KGRGRSRTPTRKGRHSRSPQWRRSRQAORGRSRSPQRGRSRSPQRPGRSRNNTOR 550
OY 471 --ESKFPYSGSYSRSSYDSSQGLAONHIALTLSPAAHTLIDLHDHHPPEAEARSA 528
Db 551 RGRSRARRGRSHSRSPAT--RGRSRTPPARGRSRSRTPARRSRSRTPARRSRSR 607
OY 529 MIVHMPDLMDIAHARSPPYR-----RYRSRGRSP-----PEFRGOS--- 566
Db 608 TPARR-----GRSRRTPARRSRTPRSPVRRSRGRSPARRGRSRTPARRGRSR 661
OY 567 -PTKRVNPEEKEREYFNRYREVPYDIKAYIGRVDPRDPE-----KEYY 613
Db 662 TPARRGRSRTPARRGRSRTP-----ARRGRSR--RTPRRGRSRSLVRRGRSH 715
OY 614 REMERKRYREYKYYGVAVGAQPPSANREDSPPELLPLNIRNSPFTTGRREDYAAQ 673
Db 716 RPPQRRGRSGSSSEKKNKSTSOR--SKNSSP-----MKKRISRRRSLSR 766
OY 674 SHR-----NRNLGNYV--EKLST--RDSHNAKNPKSKE-----KES 707
Db 767 SKAKSLSLRSLSGSSPCQKQKQTPRRSRSGSSQPKAKSRTPRRSRSSSSPPPKOK 826
OY 708 ENVPGGCKKHKHKHKKRRREEKGESESLNPELLETS-----RCRQSGSIDERTPT 762
Db 827 SKTPSRQSHSSSPHPPKVKSGTPPRQG-STTSPQANEQSVTPQRKCFESPPPELKSRT 885
OY 763 LFLVLSR-----DDATPVDEPMDAES-----ITFKSVSDKDKREKDKPKV 803
Db 886 ----PSRHSCGSSPPRVKSTTPRQSPSRSSSPQKVKALISPRQSRHSGSSSPSPSV 941
OY 804 KSDTKRKRS-----DGSATANKDNLVLPKSGPQEKVDGDKRKSRA--- 843
Db 942 TSKTTPRRSRSVSPCSNVEGRLLPRYSHSGSSPDKVKPEETPRQSHSGSISPYPKVA 1001
OY 844 -----SEPPLLKAKKEATKIDSVKPS-----SSSQKDK--VTGTPRKAS 882
Db 1002 QTPPGPSLGSKSPQEKDKSLVQSCPSLSLCAGVKSTPGESEYFCVSSLQKQOS 1061
OY 883 KSAKDRRQ--SOPRTSRKTVPTKTSQKQAPVTRPRRLKINYLIAEKNEREKRRK 941
Db 1062 QTPSDHSDTSPFVQSHSESPLOSQKQSPKGRSRSSSVTELASASPIRQDRGER 1121
OY 942 SV-----DKDESSSMKISKVEGTEIVKPSPKRMEDVEKLETPPKDIASS 990
Db 1122 SASPMLKSGMSPQSRQSSSYPTVDSNLSLQGSRLFEAAESK-EKMAIPDEQ-ATA 1178
OY 991 TTPAKKIKLN-----RETKKIGNAEASNTKPESEKLESTESKIKQ 1032
Db 1179 SPPROKDKFSPFPVQDRPSSLVFKOTLRTPPRBSGAGSSPEPKQNSALPTSSODEL 1238
OY 1033 EKYGAKKRRVAGSESSSTLVQYTTSTSGSPVRKSEKTDTKRTVITKMEYNNDMT 1092
Db 1239 MEYVERSEEPAGOLISLSELEKMTSNEESP-----EVERPAVSLILQDSQOAS 1292
OY 1093 APAEDVITIMQVQ--SKWDKDFESEEDVKTTPQIQ--SVGKPSIINKVTTKPSATA 1148
Db 1293 LEA-----VEVPSMASSWGPHFSPHKEL-SNSPLRENSFGSP--LEFRNSGPIGTE 1342

QY 1149 KYTEKESQPEK-----LQKLPKASHELMQLHPLRSSKSSASSEKQ----- 1189
DB 1343 MNTGSEVSEKEDLNGPFLQLETDPSLDKKEQSTNRS-GHSSSESPDAVEKAGMSSNOS 1401
QY 1190 -----RAKDEHSGSE-----KD-----NPKR-KSGAOP---DKESTVDRLESG 1226
DB 1402 ISSPVLADVPRTSPRSRSGSSASPEKQDLPRTPSRRSSSGSPGLRDSGTPSRRLSLG 1461
QY 1227 H-----FKTLISQ-----SSKETRTSEKHESVRSQSS-----NKDFTPGDRKKVDYD 1266
DB 1462 SSPGMKIDPRTPEFRGSECDSSPEPKALPQTPRPSRSPSSPELNNKCLTPQREKSGSES 1521
QY 1267 SRDY-----SSSKRRD-----ERGL-----ARKKDPKCKESLSG- 1298
DB 1522 SYDQKTVATPPLGQSRSGSSQELVKKPSASPOERSESDSSPDKAKTPTPLRQSRSGS 1581
QY 1299 -----QSKLR-----EERDLPKKGAKSKSNSPPRDKKPHDKAPYETRRPCE 1343
DB 1582 SPEVSKSLSPRRSRSGSSPEVKOKPRAAPRAAQSDDSPPEKAPAPRALPRRSRSGSS 1641
QY 1344 E-----TKPYDKNSGKERKHAALNKGESGANCHVYLTROTLPWRRS 1388
DB 1642 SKRGSPSPGSSSTESSPEHPKSRFTARGRSSSPPEKTKSRTPPRRS 1690
RESULT 10
QJULT1 PRELIMINARY: PRT: 4880 AA.
AC QJULT1;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MULTIDOMAIN PRESTYAPIC CYTOMATRIX PROTEIN PICCOLO.
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20170257; Pubmed=10707984;
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a presynaptic zinc finger protein structurally related to
RT Bassoon".
RL Neuron 25:203-214(2000).
DR EMBL: AF138789; AAF07822.2; .
DR HSSP: P04410; 1A25.
DR InterPro: IPR000008; C2.
DR InterPro: IPR002965; PDZ.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00168; C2; 1.
DR PRINTS: PR01217; PRICHEXTENSIN.
DR SMART: SM00239; C2; 1.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS00499; C2_DOMAIN_1; UNKNOWN_1.
DR PROSITE: PS50004; C2_DOMAIN_2; 1.
DR PROSITE: PS50106; PDZ; 1.
KW Matrix protein.
SQ SEQUENCE 4880 AA; 530140 MW; 73951E4ED83EA68 CRC64;

Query Match 5.4%; Score 396; DB 11; Length 4880;
Best Local Similarity 20.4%; Pred. No. 3.9e-12;
Matches 323; Conservative 199; Mismatches 563; Indels 502; Gaps 73;

QY 23 PTIDAAYAIKKKKRPPFLPEERSSSEDDPIRALLLCLICDINTMDAV-IPCCGN-S 80
DB 490 PTKSISQIVTGRPLDP-----PTSAQPPAQLSKTICPLCN--TTELLHPIPERANFN 542
QY 81 SCDECIKRTT-----LLESDKHTCPTCHQN-----DVSP 108
DB 543 TCEQCSYCSLCGFNPHLTETIKEMILCLNCQMORALGDLAALIPSSQPTPKAATAP 602

QY 109 DALIANKFLROAVNNFKNE-----TGYTKRLKQLPPEFLVPPPPRLSORNLQPRSRPI 164
DB 603 TATASKSPVSOQASPKKPPKQSDPKALSKRP-----EKKKPEPKPPKPPK 658
QY 165 LKQDPPVVRVTPSPCSGTAGSCSDGTLRLPAFSSILTSNOSLAPVSGN--- 221
DB 659 VQD-----PRLHG-PPIATAPQLVALEALPEP-----APPKESGL 694
QY 222 PSSAPAPVPI-----TATVISVHSEKSDP-----FRSDNKLPAALITSEHSCA 270
DB 695 PROAKAPVDVBPQKQKMETRADIQSSSTTKPDILSSQVSOAQVKATSLKTDSAKPS 754
QY 271 SSIIATLALMEEGVGTSPWNSTFVQSLHQLIPTTPRVINARPGGGRGWHENSK 330
DB 755 QGFPPPG---EK---TTPLDISKAMPRPASDSKII---SQPPG--- 788
QY 331 LGVLVSPQOIRRGERSYRSINGRHNSERSORTQSPSLPATPCFVPPPPPLVPPPH 390
DB 789 -----SESKDKHIDPIQKDEPK-----KAQPKSGKPE 818
QY 391 TLPLPPGVPPQFSPQFSSQPTAGYSVPFGFPAPANISTA-CFSPGVPTAHSTNP 449
DB 819 TKVPKGSPT-----PSGTRPTAGQAAPRQSQPKPQDSRFSGLNGITDAPKQSP 871
QY 450 TQOAPLLSREEF-----YREOND-----KGRESKFPYSG-SSISRSSTYDS-----SQGLA 494
DB 872 TTPQETVTKLFGFASISQASNLISTAGQGPHPQTPAPASKQAPPPSPAAQGPRA 931
QY 495 OIHIALTLSPSAANTLDLHDHPPEEAARSAMIVHMDLMDIAHARSPPRYRYS 554
DB 932 KSGQLPPAPAKATAV-----KKAQKAAA-----E 957
QY 555 KRSBPPEFRGQSPTKRNVPREKEREYFNRYREVPPIYIKAYGSDVDFPEKERYR 614
DB 958 NLESKPE---QAPTAKTETKDKK-----PPP---AKVGRP---PPSEBEK-- 993
QY 615 EMERKRYEMVEYKGYAAGA-----QPRPS----- 640
DB 994 -----AVAHKPKDKTTPKPKPCPLCRLTNLNGSEPPNFNTCTECKNO 1036
QY 641 -ANREDSP-----ERLLPLIRNSPPTGRREDYAAQSHRNRLNGNPEKISTRD 692
DB 1037 VCNLCGFNPTPHLTETIQEWLCLNCQORALISQGLM-----GKMPRAS--- 1081
QY 693 SHNAKDNPKSKEKSENV- GDGKGNKHKHRRNRNEKGESESF---LNPELLTSR 747
DB 1082 GPKASPMAPAPBESSQKTPGTQVKGKKKEAEKTEAEKVPVEKETASIEKTPMVTTQD 1141
QY 748 KCRGSGCIDETKTDTLFLVLP SRDQATPVVDEPDADASTIFKYS- DKDKRE---KDKPKV 803
DB 1142 KLEESG---KSKVSALPEK-----KSEEE---KALSADKRRKPPAEKPP 1185
QY 804 KSDK-----TKRSDGSATAKKDNVLY-----PSKGQPKV---GQDREKSPRSP 846
DB 1186 EEKKPLPVQKKLPPEAKPLPSSBGEKHEILKAHVQIPEEPPTGKVAAGKEEQQDSRP 1245
QY 847 PLKAKEEATKIDSVPRSSSQKDEKVTGTPRKAKSADKTRROSOPTRRSKRTVPRT 906
DB 1246 ---EALPGATPLLPK---AGEKERA VAPQAGSSKDGGER--SKETEKEDSDS 1297
QY 907 SSQKSQPVTRPRRSLRKINYL---IAREKNEEKKKSVYDKDFESSMAKISVGEETEV 963
DB 1298 SSG--QP---KSPQGLSDTGYSDGISGELPSLIPSEKDLGLKKDSDSQSSPS 1352
QY 964 KRPKRRKMGDV-----EKLETPPEKDIASSTPPAKKIKLINRETKKIGNA 1010
DB 1353 SPDLAKLSTVLSITIEAQA STLVGKAEKTKPOKI---SPEK-----P 1394
QY 1011 ENASTTRPESEKLESTSSKIKQEKVKGAKRKVAGSGSSSTLVDTYSSTGSGVPRKS 1070
DB 1395 QDQKQQTGTASETLIDIT---ISSEIKESQEKKYSPKKDSQ-----GFPSSRKE 1439

| | | | | |
|----|------|---|---|------|
| Qy | 1071 | -EEKTD----- | -TRRYIKMEEYNNNDNPAADVIMIQVQSMKMDPDSESEEDVKT | 1123 |
| Dd | 1440 | HKEPELVYDLSPPRASYSDEVEDSEENS | PVYRRKRTSTIGSS--SSDEYKQDS---- | 1493 |
| Qy | 1124 | TOPLOSVGKPSIIKNWTTTPASATARTKTESEOEPLKTLKREAS--HELMQHELRSSK | 1181 | |
| Dd | 1494 | -----QSGGEEDFIRKQIIMSDADEDASGSEDE--EFITSQKELTSVGGEQKREANCK | 1547 | |
| Qy | 1182 | GSASSEKGRAKDRHSGSEKDNPKRRSGAQDPKESTYDRLSEOGHFTLSQSSKE---- | 1237 | |
| Dd | 1548 | GKGVAGKHRRLTRKSSSTFDDDAGRHSWHDDEDE--TFDESEBELKFRFTKSOSESELVYA | 1606 | |
| Qy | 1238 | ----- | -----TRTSEKHESRGSSN | 1252 |
| Dd | 1607 | GGGGLRRFRTKTELNSTADKYSESSQKTTILYFDEPELEKESULTDSEPRSRGEGSS | 1666 | |
| Qy | 1253 | ---KDFPGRD---KKVDYDSRDYSSSKRRDGEGLARKOSPGRKESL-----SGQ | 1299 | |
| Dd | 1667 | LHASFPTGSPSPYSVSLDDSDSPSRHK-----KGSKQQRKARRHSRSGPLPTTEDSSE | 1722 | |
| Qy | 1300 | KSULREERDLPK-----KGAESKKSNSSPPRDKKPHD-----HKADYETKRRCPEETTKPYDK | 1350 | |
| Dd | 1723 | EEELREEEELTKEOQKORELEOQORSSSKSKSKODKDELRQRRRRRPPKTPPSNLPIED | 1782 | |
| Qy | 1351 | NSGKREKHAHAERANGKSESSGANCXY | 1377 | |
| Dd | 1783 | ASPTPEELROAAEM---EELHRSSCSLEY | 1806 | |

| RESIDUE | 11 | | |
|----------|---|--------------|---------------|
| Q9JRS6 | | PRELIMINARY; | PRT; 5085 AA. |
| ID | Q9JRS6 | | |
| AC | Q9JRS6; | | |
| DT | 01-OCT-2000 (TrEMBLrel. 15, Created) | | |
| DT | 01-OCT-2000 (TrEMBLrel. 15, Last sequence update) | | |
| DT | 01-DEC-2001 (TrEMBLrel. 15, Last annotation update) | | |
| DE | MULTIDOMAIN PRESYNAPTIC CYTOMATRIX PROTEIN PICCOLO. | | |
| OS | Rattus norvegicus (Rat). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Euteleota; Rodentia; Sciurognathi; Muridae; Murinae; Rattus | | |
| OX | NCBI_TaxID=10110; | | |
| RNA | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.; | | |
| RT | "Multidomain presynaptic cytomatrix protein Piccolo, long splice | | |
| RT | variant."; | | |
| RL | Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL: AF227534; AAF63196.1; -. | | |
| DR | HSSP: P04410; 1A25 | | |
| DR | InterPro: IPR000008; C2. | | |
| DR | InterPro: IPR001478; PDZ. | | |
| DR | InterPro: IPR002965; P_1rich.extensn. | | |
| DR | Pfam: PF00168; C2; 2. | | |
| DR | PRINTS: PRO1217; PRICHEXTENSN. | | |
| DR | SMART: SM00239; C2; 2. | | |
| DR | SMART: SM00228; PDZ; 1. | | |
| DR | PROSITE: PS00499; C2_DOMAIN_1; UNKNOWN_1. | | |
| DR | PROSITE: PS50004; C2_DOMAIN_2; 2. | | |
| DR | PROSITE: PS50106; PDZ; 1. | | |
| KW | Matrix protein. | | |
| SEQUENCE | 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64; | | |

```

Query Match          5 4%:  SCORE 396;  DB 11;  Length 5085;
Best Local Similarity 20.4%:  Fred. No. 4.1e-12;
Matches 323;  Conservative 199;  Mismatches 563;  Indels 502;  Gaps 73;

Qy      23  PTIDEAAYAI GKKEKPPLEPPEESSSEDDPDPAILLCILCKIDIMTAAYV-I PCGN-S 80
      ||      :  ||      ||      :  ||      :  ||      :  ||      :
Db      490  PTKSSQVYTGRLPP-----PTSAQTPADGLSKTICPLCN--TTELLHIPEKANFN 542
      ||      :  ||      :  ||      :  ||      :  ||      :
Qy      81  SCDECI RY T-----LIESDKHTTCPTCHN-----DVP 108
      ||      :  ||      :  ||      :  ||      :  ||      :

```

| | | | |
|----|------|---|-----------|
| Db | 543 | TCGCGASTVCSLGFENPNPHLTELKEKWLCLNCGOMRALGSDLAALPSSQPTPKAAATP | 602 |
| Oy | 109 | DALIANKFLROAVNNFNE-----TGITKRLKQOLPPRLFLVPPRRLSQORQPSRSPI | 164 |
| Db | 603 | TATASKSPVPSQOASPKRPPKSKODSPKALESKPP-----EPKRPPEKPPKPPPL | 658 |
| Oy | 165 | LROODPVVFRTYVSPGCSODTKTACSCSDGTLRLPAPSSISLTSNOSSILAPVSGN--- | 221 |
| Db | 659 | VKO-----PTLHG-PTATAQLVYALMLPP-----APPKESGGL | 694 |
| Oy | 222 | PSAPAPVPI-----TATVSIHVHSEKSDP-----FRDSNKLPLPAALTEHSGKA | 270 |
| Db | 695 | PEQAKAPYGDVEPQKMITETRADIOSSTTKPDILSSQOAOQVKTASPLTKDSAKPS | 754 |
| Oy | 271 | SSILFALMEKGVKPGTSPMNSIFVYGSLHGLIPTTGVRINARPGGRGWHSHSK | 330 |
| Db | 755 | QSPFPTG---EK-----TPIJDSKAMPBADSXII-----SQPGG----- | 788 |
| Oy | 331 | LGLVSPPOQIRNGERSCYNSINGRHSERSORTOSPLATPCFVPPRPLYPBPH | 390 |
| Db | 789 | -----SESKDPKHIDIQKDEPK-----KAQKGSFPE | 818 |
| Oy | 391 | TLPLRPVPPRPPQFSPORPSSOPRPAQSVPPRPPRPAALISTA-CFSGVPTAHSNTPR | 449 |
| Db | 819 | TKRPVKSPT-----PSCGTRPAGOAAPPQOAPPOBOSRFLNIGLTDARKSDP | 871 |
| Oy | 450 | TTQAPLLSREEF-----YREOND-----KRESEKFPYSG-SYSRSASYDS----- | SOGLA 494 |
| Db | 872 | TTPOETVYTGKFGGASIFGASISLITACQOGPHQOTGPAASXKAPRPSQSPAQSGA | 931 |
| Oy | 495 | QHIALTLSPRAAHTLDLHDHPREBEAARSAMTYHMDLIDIAHANSRSPYRRYS | 554 |
| Db | 932 | KSTQQLPAPAKATAV-----KKEAAMAAA-----E | 957 |
| Oy | 555 | RSRPPRPPGSGSPKRVNPREKEREYFNRYREVPYDIAKAYGSRVDFRPFEXER | 614 |
| Db | 958 | NLEKPE---OAPPAKTEKDKK-----PPP---AVYGRP---PPSEPEK--- | 993 |
| Oy | 615 | EMERKRYMEYKYYGAYAGA-----QRPFS----- | 640 |
| Db | 994 | -----AVPAHKPDKTKTKPKPACPLCRLTELNGOEPNFMTECKKQ | 1036 |
| Oy | 641 | -ANREDFSP-----BRLLPYNIRNSPFRGRREDYUAGOSHNNLGNVPEKLESTND | 692 |
| Db | 1037 | VCNLGCFPRPHLTEIOEWLCLQTORAISGLDM-----GKNRPAPS--- | 108- |
| Oy | 693 | SHNAKDNPKSKEKSENV- -GDGKGKHKHKKRRNRNEKEGSESEF---LNPDELLETSR | 747 |
| Db | 1082 | GPKASPMPAPAPESPQKTPGTGYKGGKKBAEKEKTAEBKVPKEKTAIEKTPPMVTTQ | 114- |
| Oy | 748 | KCRSSSGIDEKTDTLVLVLSRSDATPVROBPDAESTIFKXS -KDKRE-- -KDKPV | 803 |
| Db | 1142 | KLEBSEG-----KKSVALPEK-----KPEEBE---KALSANKKEKKPPAEKKPL | 118- |
| Oy | 804 | KSDK-----TKRSDSATAKKNLVK-----PSKQPOEKV---GDGREKSPRSEP | 846 |
| Db | 1186 | EKKRPIYVDKKLPPEARPLSEBGEKEKELLKANVQIDEEPRTGKAAKAGBEOQDSDNR | 124- |
| Oy | 847 | PLKAKEBAPTIDSVKSSSSQKDEKVTGTPRKAHKSASAKDTRQSQOPTRNSKRTVPNT | 906 |
| Db | 1246 | ---FALPGATPLTPK---AGEKERAAPOAQAQSSGDGGER -SKETTEKEEDSDTS | 129- |
| Oy | 907 | SSQKSOVPRTRRPSLEKINVL---JAREKNEBEKRRKSVXKOFESSMKIKVEGTEIV | 963 |
| Db | 1298 | SSQ--QP---KSPQGLSDTGTSGDGISGLGELPRLPSLSDPDKLGLKKDSFQSSPSS | 135- |
| Oy | 964 | KPSFKRMEGDV-----EKLETPKEDKIASSTPYPAKKIKINRETGKIGNA | 1010 |
| Db | 1353 | SPSOLAKLESIVLSILBAQASTVIGERAKKQPOKI---SPEK-----P | 139- |
| Oy | 1011 | ENASTPEPSEKLESTJSKQEKYVKAQKARKYVASEGSSSTLVYTTSSISGSGPVKRS | 1070 |
| Db | 1395 | ODQOQOTJASTLDTIT---ISEEIKESQKQKSPKPKDSO-----GPRSRE | 143- |

Db 934 DTKDVGATEPAD-----ETPKKIIKKTEKSDSSISQKASATDEKVSQKE--ODEPT 985
 QY 1143 KPSAT-----AKYTEKESQEPKLOKTPKASHIEMOHIELRSSKSGSSKGRAKDE 1195
 Db 966 KPAVEETQMTAEADSKSKOKETDEKL-KIDAEIAKTYQEADEKSKLDQOEIKKVSIED 1044
 QY 1196 HSGSKDNPDKKRSGAO--PDKESTVD--RLSQGHFKTLTQSSKEKRTSEKESYR---- 1248
 Db 1045 AARKEKELNDKLLESELATKASADKLEEQAOAKKAEVAAKKOKEKEDOLKLTDE 1104
 QY 1249 GSSNK-----DFTPGRD--KKVYDSRDYSSSKRDERGELARRKDSPPR 1291
 Db 1105 AASKKAAAEKLEKQAOIKKAAGADAVYKKOKELEKKNLEANKKSAAKLEKLEESAAK 1164
 QY 1292 GKESLSGQ-----KSKIRER-DLPKGAESKSSNSPPDKPR-----HDHKAPY 1336
 Db 1165 SKQTEYEOAKLDAQTKAKTAKEQKLEKDEKSTKESKESKEVDEKPKKKVLTAKTEKSDS 1224
 QY 1337 ETKRCEETKPYDKNSG--KEREKHAAEARNGKES 1369
 Db 1225 SISQKSETSKYVESAGPSESETQVADAARAKOKE 1260

RESULT 14

Q9UHA8 PRELIMINARY: PRT: 2296 AA.
 AC Q9UHA8: 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 13, Last annotation update)
 DE SPLICING COACTIVATOR SUBUNIT SRM300.
 GN SRM300.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N. A.
 RX MEDLINE=20132238; PubMed=10668804;
 RA Blencowe B.J., Bauren G., Eldridge A.G., Issner R., Nickerson J.A.,
 RA Rosonina E., Sharp P.A.;
 RT "The SRM160/300 splicing coactivator subunits.*";
 RL RNA 6:111-120(2000).
 DR EMBL: AF201422; AAF21439.1; -;
 DR InterPro: IPR002965; P_Trich_extensions.
 DR PRINTS: PR01217; PRICEXTENS.
 SO SEQUENCE 2296 AA; 251964 MW; 17C0BDAEAI0A9CF9 CRC64;

Query Match 5.1%; Score 377.5; DB 4; Length 2296;
 Best Local Similarity 19.9%; Pred. No. 115e-11;
 Matches 356; Conservative 199; Mismatches 654; Indels 583; Gaps 75;

QY 31 AIGKEKEPFLPEEPSSSEEDDPIPAELLCICKDINTDAVYIPCCGNSCDE----- 84
 Db 48 ALVKRPNPIDLHERKR-----VELRCLLEEMME-----QGYVEEQIQEK 90
 QY 85 -CITTLLESKHTCPTCHQNDVSPDALIANKFLQAV-----NNFNKNGCYTKR 133
 Db 91 VATFRIMLE-----KDVNPGKKEETPGQAPVETETHQALAELEKKE-----R 134
 QY 134 LR-----KOLPFLFLVPPRPISQNRNLQGRSRSPILRQOD 169
 Db 135 LAAAGISDSIYDGSFPDQRRAREKQAP-----EPKPYSLVRESNNS----- 182
 QY 170 PVVFRYVPTGSDRTAGSCDSGTLRL-----PAPSISLVSNOGSLAPVYGNS 223
 Db 183 -----QPGRRRRKRRIEDAGORALLDGRERKAQRSTGQNL-----PR 225
 QY 224 SAPAPYDITTAIVSIVHSEKSDGPPRSDNKLPAALITSEH-----SKGASSIATIALM 279
 Db 226 NVSIGILPIQRAVNVNLTKE-----SGLEVOHQPPRAAGPTVQLITLIL 269

QY 280 EEKGVPTSPMNSIFVGOSLIHGOLI--PTTGPV-----NINARPGGGR 323
 Db 270 PPIPLA-----VGLEVLQKLILQPLWGEVLPLQGDAGREMLSVNOVLSTQR- 320
 QY 324 GWEHNSKGLGYVSPQOIRGERSCYRSINRGHHSERSQRTQSPSLPATPCFVVPVPP 383
 Db 321 -----ASSPETAKQSPSPEDKDKDKKESATRPSPSPERSSTGEPAPVPL 369
 QY 384 L-----YPPPHILPL--PCGVPPPOFS-----POFPSSOPPLAGISVPPGPPAP 428
 Db 370 IAEHGGSPQPLATPLSQEPVNPPEASPTDRSPPKSEKLPQSSSESSPPS--PQP 427
 QY 429 ANIST-ACSPGPV-----THASNT----- 447
 Db 428 TVYSRHASSPSPAPAPAGSHREISSPTSKNSHGAKDKSHSHPSRMGRNSP 487
 QY 448 -----APTOAPLISREEFYREOND-----KGR----- 470
 Db 488 ATAAGRSRSPPTKRGHSRSRSPQWRGRSAQRMGRSRSPQRRGRSRSPORPGMSRN 547
 QY 471 -----ESKFPYGSSTSRNSTYDSSOGLAQHIIHALTSSAAHTDLHDHPPEEAEA 525
 Db 548 TQRRGRSRGARRGRSHSRSPAT--RGRSRSTPPARRGRSRSTPPARRSRSTPPRRS 604
 QY 526 RSAMLVHMPDLMDIAHARSPPYR-----RYRSRGRSP-----PEFRGOS 566
 Db 605 RSTTPARR-----GSRKSTPPARRSRSTPSYRRSRSRSPARRGRSRSTPPARRGRS 658
 QY 567 -----PTKRVNPREEKREYFNRYREVPPYDIKAYYGRSVDFRDPPE-----K 610
 Db 659 RSTTPARRGRSRSTPPARRGRSRSTP--ARRGRSR--RTPPRRGRSRSLVRGR 712
 QY 611 ERYREMERKYREMYKYGYAVGAOPRPSANRDEFSPELLPLINRNSPPTKRGREDYA 670
 Db 713 SHSRTPQRRGRSGSSSERNNKSRSTQGR--SRNSNPE-----MKKSRISSRRSRSL 763
 QY 671 AGOSHR-----NRNLGNVP--EKLS--RDSHAKNDPKSKE----- 704
 Db 764 SPBSKAKSLSLRSLSGSPCKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 823
 QY 705 KESENVPGDGKGNKHKHKKRRNEKEGESESLNPELETS-----RKCRGSSGIDETK 759
 Db 824 KQKSKTPSQSHSSSPHHKVKVSGTPRQC-SITSPQANQGVSTPQRRSCFPSSPPELK 882
 QY 760 TDTLFLVPSR-----DDATPVDEPMDAES-----ITFKSVSDKDKREK 800
 Db 883 SRT-----PSRHSCSGSSPPRVKSTPPRQSPSRSSPPQPKVVAIISPRORSHSGSSPSP 938
 QY 801 PKVSKDTKRKS-----DGSATLAKKDVNLKPSKGPQKVDGDRKRSR 843
 Db 939 SKVTSKTTPKRSRVSPCSNVSRLLPKYSHSGSSSPDKVAKPETPPKQSHSGSISPYRK 998
 QY 844 -----SEPLLKAKEEATKIDSVPS-----SSQKDEK--YTGTPRK 879
 Db 999 VKAQTPPGSLGSKSPQCKSKSDLVQSCPSLILCAGVKSSTPPGSGSYGVSSLOLK 1058
 QY 880 AHSKSAKDRRQ--SQPTRRSKRTYVKTSSQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 938
 Db 1059 GOSQTSPPDHRSDTSSPEVQSHSESPISQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1118
 QY 939 RKKSU-----DKFESSMKISKVEGETIYKPPKRMGEGDVLEKLETPKDKT 987
 Db 1119 GEFSSAPMLKSGMSPQSFQSDSSSYPIYVDSNLSLQSRLETAESK-EKMLPQED-- 1175
 QY 988 ASSTPPAKKIKLN-----RETGKIGNAENASTKPESEKLESTSK 1029
 Db 1176 ATASPPRQKDFSPFVQDRESSLVFKDTLRTPPRERGAGASSPPTKQNSALPTSSOD 1235
 QY 1030 IKQEKYKGAKKRVAGSEBSSSTLVNDYTSSTSGSGSPVAKSEKTDTKTKVTKTMEYNN 1089
 Db 1236 EELMEVEKSEBPAGQILSHLSSELEKEMSTSNESP-----EYERAVASLTLDQOS 1289
 QY 1090 DMTAPAEVITIMIQVQ--SKMDKDFESEEDVKTQPIQ--SVQKPSIINKVYTKPS 1145

```
Db 1290 QASLEA-----VEVPMASWGGPHFSPEHKEI-LSNSPLRENSFGSP---LEFRNSGPL 1339
Qy 1146 ATAYTEKESQPEK-----LOKLPKEASHELMOHELRRSSGASSPKGG----- 1189
Db 1340 GTEKNTGSSSEVKEDLNGPFLNQLETDPISLDMKEQSTSS--GHSSSELSPDAVEKAGSS 1398
Qy 1190 -----RAKREHSGSE-----KD---NPKDR-KSGAQP---DKESTVDRLS 1223
Db 1399 NQSTSSPYLDAVPRTPPSRERSSASSPKMGDLPRTPSRRSRSGSSPGLRSGSTPSRHS 1458
Qy 1224 EQGF-----KTLSSSKTTRTSEKHESYRG-----SNKDTFGRODKV 1263
Db 1459 LSGSSPGMKDIPRTPSRGRSECDSSPEPKALPQTPRPSRSPSSPELNKLTQORESG 1518
Qy 1264 DYDSRDY-----SSSKRRD-----ERGEL-----ARKRDPPEKESL 1296
Db 1519 SESSVDQTYVARTPLGORSRSGSQELDYKPSASPOERSESDSPDKAKTRTLROKSR 1578
Qy 1297 SG-----QSKLR-----EERDLPKGAESKSSNSPPRDKKPHDKAPYETKR 1340
Db 1579 SGSSPEVDSKSRSLPRRSRSGSSPEVKDPRAPRAQSGSDSSPEPKAPAPALPRRSRS 1638
Qy 1341 PCEE-----TKPVDKNSGKERKHAAEARNGKESGANCHYLTROTLPWRRS 1388
Db 1639 GSSSKRGSPSESGSSSTESSPEHPKSRRTARCGSRSSPEPKTKSRTPRRRS 1690

RESULT 15
Q95Y2M2 ID ID Q95Y2M2 PRELIMINARY; PRT; 17352 AA.
AC Q95Y2M2
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE I-CONNECTIN.
GN I-CON.
OS Procambatus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacidea; Cambaridae; Procambatus.
OX NCBI_TaxID=6728;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=21423462; Pubmed=11532946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RT Sun P., Matsuyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RT sarcomeres of crayfish claw muscle.";
RL EMBL: AB055861; BAB64297.1; -.
DR EMBL: 17352 AA; 1962348 MW; 4BA157BEC042EA2D CRC64;
SQ
Query Match 5.1%; Score 375; DB 5; Length 17352;
Best Local Similarity 18.9%; Pred. No. 2.2e-10;
Matches 291; Conservative 180; Mismatches 602; Indels 466; Gaps 56;
```

```
Qy 241 HSEKSDGPFPRDSQDKLLPAAALTSSEHSGASSIATLMEKGVGTSPWNSLFGOSL 300
Db 6061 E-----PFKEVDSDLDDKQVQLE-----EPVAPBEKLEK--PSTPEGRPI----- 6100
Qy 301 HGOLIPPTGPRVINAARPGGRCGWEHNSKLGVLVSPQOIRREKRCYCSINGRHSE 360
Db 6101 -----PWEEKLEEGKPRPKKERT-----KPEKEESELPSW-----RGRLLP 6140
Qy 361 RSQRTQSPSLPATPCFVAVPPPLYPPEPPT--LPLPGVPPPOFSQPPSOPPTAGYS 418
Db 6141 KEEKEVQOL--KPEKPKPTK--KPKPKHKPGTTPPELPEPKTLEPYDKKMQ--- 6194
Qy 419 VPPGPPAPANISTACFSQVPTAHSNTMPTQAPLLSREEFYREQD----- 467
Db 6195 -ITPDHBPQ-----VEDEDEKLPDSBQPKLKKAVPKKDEEBEILPSWGRK 6242
Qy 468 -----KGR-----SKFYSSGSYSRSYTD 488
Db 6243 LPPKEDKEEITLTKPFKKVPRKPEKPSKLPKGPPEPEIPEPKSLPEYSKPEKESD 6302
Qy 489 SSQGLAQHIALTLSPSAHTLDLHDHPHPPEAEARSMIYMPDLMIDAHRSRSP 548
Db 6303 KIPELAKESKPEPEKPEBEPEKLDSDKKPDESEI---ITQK----- 6344
Qy 549 YRRYRSRSRSPPEFRGSP-----TKRNVPEEKEREYFNRYRVP----- 589
Db 6345 -KRRLOKLKAPTEKEFELPVTLAKTSQKVFVPEEVLLEVELEHVTTPEVEPEVEKRV 6403
Qy 590 --PPYDIKAYYGRSVDFRDPPEKERYREWEKRYREWEYKYGAVGAQPPRPSANREDFS 647
Db 6404 WSPPEYETVYVEEIPKEPEVELEKEYE----- 6433
Qy 648 PERLLPLINSPPTRRRRRDYAAAGSHRRNRNLCGNYPEKLTSHDNANONPKESKES 707
Db 6434 -----PPTKPKDEE-----EDGKYEK-----PKDPEEDRK 6464
Qy 708 ENYVGDGK-----NKKHKHRRNEKEGESESFNLPELLETSRKRGSSGIDET 758
Db 6465 LKL---GKRLRPEEKEKKKPKRRRPSKEAE--KPOLKPIPK-----KPEEE 6513
Qy 759 KTDTLFVLPSSBDATPYRD-EPMDAESI--TFKSVSKDKREKDKPVKSDKTKRKSDGS 815
Db 6514 KKDVTLPKPGKSKKIPDEEPVLEPPEPTEPILDKDVLEPKLPKPEPEKPSIE 6573
Qy 816 ATAKKQVNLPSKSGPOQKVGDDREKSPSE-----PLTKAKEAT--KIDS 860
Db 6574 PEAKPLPEKPSSEEBEKLKERPKPEKEBEAEVPSWGRKRLPPEKEKEEYLAKFPK 6633
Qy 861 VKPSSSQKDEKVTGTPRKAHNSKAKDTRQOSOPRTRRSKRYVPTSSQSOPVTRRRPR 920
Db 6634 EKPEEPKPKYKKGKGYEPEIPEPEKT--PLBYTTPDKKEKVDGTYPEYKPEDEKPK 6691
Qy 921 SLRKI---NTLIAREKNE-----REKRSYVDKDESSSMKISKYGEIYKPSPRK 970
Db 6692 PEEBIRKPKBRKIKDKDEEVETPSWRGKRLPPEKEDEEITLTKPFKKEKKEPSPKIK 6751
Qy 971 -----MEGOVEKLE--RTPEKDKIASSTPAKIKILNRELGKIKGNAENASTTKEP 1019
Db 6752 PGKPYEPIPEPEKPRPELEPEKPEKRVDPQVPRKEKPEVQ-----PEPEKEI 6802
Qy 1020 SEKLESTSSKIKQKVGKARKKAVAGSESSSTLVDTYSTSGSGVPRKSEKTDTRK 1079
Db 6803 SPKPEKEKPOOLEKIPGKKRR-----PRMIAAI---EKFEIPIK 6842
Qy 1080 VIKTMEYINDNTAPADVITIMIQVPSKDKDDESEEDVDKTTQIQSVGKPSIKN 1139
Db 6843 TLKK-----TQNVFVPEEVLLE-----TIELHVEVPEVIE-----PEVEKR 6880
Qy 1140 VTIKPSATAYT-----EKSEQPEKLOKL-----PKEASHELMOHELRRSSGASSE-- 1187
Db 6881 WSPPEYETVYVEEIPKEPEVELEKEYEPPTKPKDEEBEEDKGYKERRPKDKPEEED 6940
```

Fri Sep 13 10:04:45 2002

us-09-811-045a-1.rsp

Page 15

```

Oy 1188 -----KGRAKDDEHSGSEKNDNDKRRKSGAORDDKSTYDRLEQCHFTLSGSSKEKRTS 1241
Db 6941 RKLKLGKGLKRPBEEGEEKLAPPKRRP-KSPEKE-----AEKQOLKPIPKKRPBEEK 6993

Oy 1242 EKHESVSGSSNKDFTPBQDKVYDSDRDYSSSKRRD-EKGEIARRKDDSPPKGESLS-- 1297
Db 6994 DKVTPKRGKKRPPKKIPDRP-VPLEPFEKTEPELIDKQVPLEKTLAKPKPPKKPSPLEP 7052

Oy 1298 -----GOKSLKREERDLPPKKAES-----KKS 1319
Db 7053 EAPKPLEEKSEEBEERPLKKEKRRPKPEKKEEBEEDVPSWGRKRLPPKPEEKKEIVLPKPEKE 7112

Oy 1320 NSSPRDCKRPHDKAPRTERRPCEERKPYDKNKSGKEREK 1358
Db 7113 KPEERKPKPKYKRGKRPLEPRPTEPRPTEPRPTEPRPTEPRPTEPRPTEPRPTEPRPTEPR 7151

```

```
Search completed: September 12, 2002, 19:12:09
Job time: 10377 sec
```

THIS PAGE BLANK (USPTO)